

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 08:59:02 ; Search time 1931 Seconds

(without alignments)
14378.082 Million cell updates/sec

Title: US-09-880-729A-1

Perfect score: 954
Sequence: 1 atgggtgtgtatccttttga.....gagagatagcatgataaa 954

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_com:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htg_hum:*
40: em_htg_mus:*
41: em_htg_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	954	100.0	954	6	AR077712	AR077712 Sequence
2	954	100.0	954	6	AR096552	AR096552 Sequence
3	954	100.0	954	6	AR157661	AR157661 Sequence
4	952.4	99.8	10244	1	AE001813	AE001813 Sequence
5	235.8	24.7	1083	6	AR205090	AR205090 Sequence
6	153.2	16.1	2160	1	CLOCCELCDD	D90341 C. celluly
7	151.6	15.9	3002	1	CLOCCEH	M31903 C. thermocel
8	82.2	8.6	10415	1	AE007598	AE007598 Clostridi
9	78.6	8.2	1756	8	NPENDGUB	Z31364 N. patriciar
10	76	8.0	3471	1	CLOCCEH	M22759 Clostridium
11	75.4	7.9	1451	8	AF015249	AF015249 Orpinomyces
12	74.8	7.8	1876	8	AF165266	AF165266 Pitomyces
13	73	7.7	1825	8	OSU57818	U57818 Orpinomyces
14	73	7.7	1826	6	AR130363	AR130363 Sequence
15	72.6	7.6	5502	8	PE0277483	AJ277483 Pitomyces
16	71.8	7.5	1998	1	CLOCENB	M75706 Clostridium
17	68.6	7.2	1532	6	AR108136	AR108136 Sequence
18	68.6	7.2	1532	8	U97153	U97153 Orpinomyces
19	68.6	7.2	1966	8	AF015248	AF015248 Orpinomyces
20	63.2	6.6	3949	8	AF053363	AF053363 Neocallim
21	62.8	6.6	1086	8	AF094757	AF094757 Pitomyces
22	60.4	6.3	7218	6	I66494	I66494 Sequence 14
23	58.6	6.1	2000	1	CLOCENGD	M37434 Clostridium
24	58.4	6.1	1634	8	NEU38843	U38843 Neocallim
25	58.2	6.1	10705	1	AE007666	AE007666 Clostridi
26	53.8	5.6	10894	1	AE007844	AE007844 Clostridi
27	53	5.6	247699	2	AC106241	AC106241 Rattus no
28	50.6	5.3	1550	1	RUMCELNA	L10243 Ruminooccu
29	50	5.2	163208	2	AC124923	AC124923 Rattus no
30	50	5.2	300750	1	AP001509	AP001509 Bacillus
31	47.6	5.0	149834	2	AC105683	AC105683 Rattus no
32	47.4	5.0	146383	2	AC116367	AC116367 Oryza sat
33	46.8	4.9	166096	2	AC120947	AC120947 Rattus no
34	46.6	4.9	1141	6	AX083744	AX083744 Sequence
35	45.6	4.8	64688	2	AC131425	AC131425 Rattus no
36	45.6	4.8	161844	2	AC106526	AC106526 Rattus no
37	45.2	4.7	125020	9	AF429315	AF429315 Homo sapi
38	45.2	4.7	200820	2	AC129459	AC129459 Rattus no
39	45	4.7	122737	2	AC109901	AC109901 Rattus no
40	45	4.7	124524	9	AC117411	AC117411 Homo sapi
41	44.8	4.7	774	11	CNS06FFX	AL396355 T3 end of
42	44.8	4.7	58696	2	AC103749	AC103749 Homo sapi
43	44.8	4.7	161389	10	AL691440	AL691440 Mouse DNA
44	44.8	4.7	171462	2	AC117832	AC117832 Homo sapi
45	44.8	4.7	207804	2	AC011714	AC011714 Homo sapi

ALIGNMENTS

RESULT 1
AR077712
LOCUS AR077712 954 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 1 from patent US 5962258.
ACCESSION AR077712
VERSION AR077712.1 GI:10004458
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 954)
AUTHORS Mathur, E.J. and Lam, D.E.
TITLE Carboxymethyl cellulase from *Thermotoga maritima*
JOURNAL Patent: US 5962258-A 1 05-OCT-1999;
FEATURES Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a

source	1. .954	/organism="unknown"
BASE COUNT	335 a	145 c 221 g 253 t
ORIGIN		
Query Match	100.0%	Score 954; DB 6; Length 954;
Best Local Similarity	100.0%	Pred. No. 3.2e-203;
Matches	954; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1	ATGGGCTTGATCCCTTTGAAAGGAACAATAATTTGGAGAGCGCTTATATAGGAAT 60
DB	1	ATGGGCTTGATCCCTTTGAAAGGAACAATAATTTGGAGAGCGCTTATATAGGAAT 60
QY	61	GGCGTTGAAGCACCAATATGAGGAGAGCTGGGAGTGTGATATAAGATGATTTCTTGAC 120
DB	61	GGCGTTGAAGCACCAATATGAGGAGAGCTGGGAGTGTGATATAAGATGATTTCTTGAC 120
QY	121	ATTATTAAGAAGACCGGTTCTCTCTGTTTGCATTCGATATAGATGAGTACGACGCT 180
DB	121	ATTATTAAGAAGACCGGTTCTCTCTGTTTGCATTCGATATAGATGAGTACGACGCT 180
QY	181	TACGCGTTTCTCTCTTATTAATCATGATCGCTCTTCAAAAAGTGGATGATGAT 240
DB	181	TACGCGTTTCTCTCTTATTAATCATGATCGCTCTTCAAAAAGTGGATGATGAT 240
QY	241	AACGAGACCCGTGAAGAAGAGAGCTGGCTGGTATTAATTTATCATCACTCGAGAGTTA 300
DB	241	AACGAGACCCGTGAAGAAGAGAGCTGGCTGGTATTAATTTATCATCACTCGAGAGTTA 300
QY	301	ATGAATGATCCAGAGAAACAACAAGAAAGATTTCTTGCTCTTTGGAAACAATTCCTGAT 360
DB	301	ATGAATGATCCAGAGAAACAACAAGAAAGATTTCTTGCTCTTTGGAAACAATTCCTGAT 360
QY	361	CGTTATTAACATATCCCGAACACTATTTTGAATTTGTGATGAACTCAGCAAAAT 420
DB	361	CGTTATTAACATATCCCGAACACTATTTTGAATTTGTGATGAACTCAGCAAAAT 420
QY	421	CTTACTCCGGAANAATGGAATGAACTGCTTGAGGAACCTCTAAAAGTTATPAATCAATT 480
DB	421	CTTACTCCGGAANAATGGAATGAACTGCTTGAGGAACCTCTAAAAGTTATPAATCAATT 480
QY	481	GACAAAAGCACTATTAATTTATGACAGACGCTAATGGGGGGTATATCGCCCTTGAA 540
DB	481	GACAAAAGCACTATTAATTTATGACAGACGCTAATGGGGGGTATATCTGCCCTTGAA 540
QY	541	AAACTGTCTGTCCCAAAATGGGAAAAAATTCATAGTTACATTCACCTACATCTCT 600
DB	541	AAACTGTCTGTCCCAAAATGGGAAAAAATTCATAGTTACATTCACCTACATCTCT 600
QY	601	TTGCAATTTACCCATCAAGAGCTGAGTGGTGGAGAGATCTGGAANAATGGTTGGAGAA 660
DB	601	TTGCAATTTACCCATCAAGAGCTGAGTGGTGGAGAGATCTGGAANAATGGTTGGAGAA 660
QY	661	AAGTGGGAGATCTCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGAGAA 720
DB	661	AAGTGGGAGATCTCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGAGAA 720
QY	721	TGCTCAAAAAAGAACAAAAAGCAATTTACATAGCTGAGTGGTGGCTACAGAAAAAGCT 780
DB	721	TGCTCAAAAAAGAACAAAAAGCAATTTACATAGCTGAGTGGTGGCTACAGAAAAAGCT 780
QY	781	GACCTTGATACAGATTAATAATGAGACCTCTTGTGCTTCGGAATGAGAAAAAGAGA 840
DB	781	GACCTTGATACAGATTAATAATGAGACCTCTTGTGCTTCGGAATGAGAAAAAGAGA 840
QY	841	TGAGCTGGGCACTACTGGGAATTTTGTCCGGTTTGGTATTGATACCTCGAGAAA 900
DB	841	TGAGCTGGGCACTACTGGGAATTTTGTCCGGTTTGGTATTGATACCTCGAGAAA 900
QY	901	ACCTGGAATTAAGATCTTTGAACTTAAATAGAGAGATAGCATTTGAATTA 954
DB	901	ACCTGGAATTAAGATCTTTGAACTTAAATAGAGAGATAGCATTTGAATTA 954

LOCUS	AR096552	954 bp	DNA	linear	PAT 08-SEP-2000
DEFINITION	Sequence 1 from patent US 6008032.				
ACCESSION	AR096552				
VERSION	AR096552.1	GI:10025457			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 954)				
AUTHORS	Mathur,E.J. and Lam,D.E.				
TITLE	Carboxymethyl cellulase from <i>Thermotoga maritima</i>				
JOURNAL	Patent: US 6008032-A 1 28-DEC-1999;				
FEATURES	Location/Qualifiers				
source	1..954				
BASE COUNT	335 a 145 c 221 g 253 t				
ORIGIN					
Query Match	100.0%; Score 954; DB 6; Length 954;				
Best Local Similarity	100.0%; Pred.No. 3.2e-203;				
Matches	954; Conservative 0; Mismatches 0; Indels 0; Gaps 0				
Db	1 ATGGGTGTGATCCTTTTGAAGAACAATAATTTGGAAGAGCATTAATATAGAAAT 60				
	1 ATGGGCTGTGATCCTTTTGAAGAACAATAATTTGGAAGAGCATTAATATAGAAAT 60				
Qy	61 GCGCTTGAAGACCAAAATGAGGAGACTGGGAGTGTATAAAGATGAGTCTTGGAC 120				
Db	61 GCGCTTGAAGACCAAAATGAGGAGACTGGGAGTGTATAAAGATGAGTCTTGGAC 120				
Qy	121 ATTATTAAGAAGACCGGTTCTCATAGTTCGAATTCATTAAGATGGAGTACACAGCT 180				
Db	121 ATTATTAAGAAGACCGGTTCTCATAGTTCGAATTCATTAAGATGGAGTACACAGCT 180				
Qy	181 TAGCGCTTCCCTCTTATTAATAATCATGATGCGTCTTCAAAAGAGTGAATGAGTATA 240				
Db	181 TAGCGCTTCCCTCTTATTAATAATCATGATGCGTCTTCAAAAGAGTGAATGAGTATA 240				
Qy	241 AAGGAGCCCTGAAGAAGAGGACTGGCTGTGCTATTAATATCATCATGACAGAGATTA 300				
Db	241 AAGGAGCCCTGAAGAAGAGGACTGGCTGTGCTATTAATATCATCATGACAGAGATTA 300				
Qy	301 ATGCAATGATCCAGAGAACAAGAAAGAAATTTCTGCTCTTTGGAAACAATTCGTGAT 360				
Db	301 ATGCAATGATCCAGAGAACAAGAAAGAAATTTCTGCTCTTTGGAAACAATTCGTGAT 360				
Qy	361 CGTTATTAAGACTATCCCGAAACTCTATTTTTTGAATTTCTGAATGAATCCACAGGAAT 420				
Db	361 CGTTATTAAGACTATCCCGAAACTCTATTTTTTGAATTTCTGAATGAATCCACAGGAAT 420				
Qy	421 CTGACTCCGGAAAAATGGAATGAAGCTGTGAGGAAGCTCTAAAGTTTAAAGATCAATT 480				
Db	421 CTGACTCCGGAAAAATGGAATGAAGCTGTGAGGAAGCTCTAAAGTTTAAAGATCAATT 480				
Qy	481 GACAAAAAGACACTATATTAATATAGGCACACCTGATGGGGGGGTATATCTGCCCTTGA 540				
Db	481 GACAAAAAGACACTATATTAATATAGGCACACCTGATGGGGGGGTATATCTGCCCTTGA 540				
Qy	541 AAACCTGTCTGTCCCAAAATGGGAAAAAAATTCCTATCTTCATTCACATACATCAATCCT 600				
Db	541 AAACCTGTCTGTCCCAAAATGGGAAAAAAATTCCTATCTTCATTCACATACATCAATCCT 600				
Qy	601 TTGCAATTTACCATATGAAGAGCTAGTGGGGAAGAGATCTGGAATGCTTTGGAGAGA 660				
Db	601 TTGCAATTTACCATATGAAGAGCTAGTGGGGAAGAGATCTGGAATGCTTTGGAGAGA 660				
Qy	661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAGAAATTCATTTTATAGAAGAA 720				
Db	661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAGAAATTCATTTTATAGAAGAA 720				

QY 721 TGGTCAAAAAGAACAAACCAATTACATAGTGTAGTTGGTCCCTACAGAAAAAGCT 780
 |||
 Db 721 TGGTCAAAAAGAACAAACCAATTACATAGTGTAGTTGGTCCCTACAGAAAAAGCT 780
 QY 781 GACCTTGAAATCAAGATTAATAATGAGCTCTTGTGTGCGCAATGGAAGAAAGAGAGA 840
 |||
 Db 781 GACCTTGAAATCAAGATTAATAATGAGCTCTTGTGTGCGCAATGGAAGAAAGAGAGA 840
 QY 841 TGGAGCTGGGCATCTGAGCAATTTGTCGGCTTTGGTGTATGATCTCTGAGAAAA 900
 |||
 Db 841 TGGAGCTGGGCATCTGAGCAATTTGTCGGCTTTGGTGTATGATCTCTGAGAAAA 900
 QY 901 ACCTGGAATAAAGATCTTTAGAGCTTTAATAGAGAGAGATGACATTTGAATAA 954
 |||
 Db 901 ACCTGGAATAAAGATCTTTAGAGCTTTAATAGAGAGAGATGACATTTGAATAA 954

RESULT 3
 ARL157661
 LOCUS ARL157661 954 bp DNA linear PART 17-OCT-2001
 DEFINITION Sequence 1 from patent US 6245547.
 ACCESSION ARL157661
 VERSION ARL157661.1 GI:16218624
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 954)
 AUTHORS Mathur,E.J. and Lam,D.E.
 TITLE Carboxymethyl cellulose from thermotoga maritima
 JOURNAL Patent: US 6245547-A 1 12-JUN-2001;
 FEATURES
 source 1..954
 /organism="unknown"

BASE COUNT 335 a 145 c 221 g 253 t
 ORIGIN
 Query Match 100.0%; Score 954; DB 6; Length 954;
 Best Local Similarity 100.0%; Pred. No. 3.2e-203;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCTGTGATCCTTTTGAAGAACAAATATTTGGAGAGGACATTAATAGCAAT 60
 |||
 Db 1 ATGGGCTGTGATCCTTTTGAAGAACAAATATTTGGAGAGGACATTAATAGCAAT 60
 QY 61 GCGCTTGAAGACCAATGAGGAGACTGGGAGTGTGATTAAGATAGTCTTCGAC 120
 |||
 Db 61 GCGCTTGAAGACCAATGAGGAGACTGGGAGTGTGATTAAGATAGTCTTCGAC 120
 QY 121 ATTATTAAGAGAGCGGTTTCTCTCATGTTTCAATTCGAATAGATGAGTACGACGCT 180
 |||
 Db 121 ATTATTAAGAGAGCGGTTTCTCTCATGTTTCAATTCGAATAGATGAGTACGACGCT 180
 QY 181 TACGCTTTCCTCTTATTAATATCATGATCGCTTCTTCAAAAAGAGTGAAGTGATA 240
 |||
 Db 181 TACGCTTTCCTCTTATTAATATCATGATCGCTTCTTCAAAAAGAGTGAAGTGATA 240
 QY 241 AACGAGACCCCTGAAAAGAGAGACTGGCTGTATTAATATTCATACATACAGAGAGTTA 300
 |||
 Db 241 AACGAGACCCCTGAAAAGAGAGACTGGCTGTATTAATATTCATACATACAGAGAGTTA 300
 QY 301 ATGAATGATCCAGAGAACACACAGAAAGATTTCTGCTTTTGAAGCAAAATGCTGAT 360
 |||
 Db 301 ATGAATGATCCAGAGAACACACAGAAAGATTTCTGCTTTTGAAGCAAAATGCTGAT 360
 QY 361 CGTTATAAAGACTATCCGCAAACTCTATTTTGAATTTCTGATAGTACCTCAGCAAT 420
 |||
 Db 361 CGTTATAAAGACTATCCGCAAACTCTATTTTGAATTTCTGATAGTACCTCAGCAAT 420
 QY 421 CTTACTCCGGAATAATGATGAAGTCTTGAAGAGCTCTAAGTATTAAGATCAATT 480
 |||
 Db 421 CTTACTCCGGAATAATGATGAAGTCTTGAAGAGCTCTAAGTATTAAGATCAATT 480

QY 481 GACAAAAAGCACATTAATTAATAGGCACAGTGAATGGGGGGTATATTCGCCCTTGA 540
 |||
 Db 481 GACAAAAAGCACATTAATTAATAGGCACAGTGAATGGGGGGTATATTCGCCCTTGA 540
 QY 541 AAATGCTCTCTCCCAAAATGGAAGAAATTTCTATGACTTCAATTCATCTACATCT 600
 |||
 Db 541 AAATGCTCTCTCCCAAAATGGAAGAAATTTCTATGTTCAATTCATCTACATCT 600
 QY 601 TTGCAATTTCCCATCAAGAGAGTGGTGGTGAAGGATCTGAGAAATGGTGGGAGA 660
 |||
 Db 601 TTGCAATTTCCCATCAAGAGAGTGGTGGTGAAGGATCTGAGAAATGGTGGGAGA 660
 QY 661 AACTGGGATCTCCAGATGATGATCAAGAACATTTGATAGAAATTCATTTTATAGAGAA 720
 |||
 Db 661 AACTGGGATCTCCAGATGATGATCAAGAACATTTGATAGAAATTCATTTTATAGAGAA 720
 QY 721 TGGTCAAAAAGAACAAACCAATTACATAGTGTAGTTGGTCCCTACAGAAAAAGCT 780
 |||
 Db 721 TGGTCAAAAAGAACAAACCAATTACATAGTGTAGTTGGTCCCTACAGAAAAAGCT 780
 QY 781 GACCTTGAAATCAAGATTAATAATGAGCTCTTGTGTGCGCAATGGAAGAAAGAGAGA 840
 |||
 Db 781 GACCTTGAAATCAAGATTAATAATGAGCTCTTGTGTGCGCAATGGAAGAAAGAGAGA 840
 QY 841 TGGAGCTGGGCATCTGAGCAATTTGTCGGCTTTGGTGTATGATCTCTGAGAAAA 900
 |||
 Db 841 TGGAGCTGGGCATCTGAGCAATTTGTCGGCTTTGGTGTATGATCTCTGAGAAAA 900
 QY 901 ACCTGGAATAAAGATCTTTAGAGCTTTAATAGAGAGAGATGACATTTGAATAA 954
 |||
 Db 901 ACCTGGAATAAAGATCTTTAGAGCTTTAATAGAGAGAGATGACATTTGAATAA 954

RESULT 4
 AE001813
 LOCUS AE001813 10244 bp DNA linear BCT 02-JUN-1999
 DEFINITION Thermotoga maritima section 125 of 136 of the complete genome.
 ACCESSION AE001813 AE000512
 VERSION AE001813.1 GI:4982321
 KEYWORDS
 SOURCE Thermotoga maritima.
 ORGANISM Thermotoga maritima.
 REFERENCE 1 (bases 1 to 10244)
 AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwynn,M.L., Dodson,R.J.,
 Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
 McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
 Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
 Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
 Eisen,J.A., Fraser,C.M. et al.
 TITLE Evidence for lateral gene transfer between Archaea and bacteria
 from genome sequence of Thermotoga maritima
 JOURNAL Nature 399 (6734), 323-329 (1999)
 MEDLINE 99287316
 PUBMED 10360571
 REFERENCE 2 (bases 1 to 10244)
 AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwynn,M.L., Dodson,R.J.,
 Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
 McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
 Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
 Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
 White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
 TITLE Direct Submission
 JOURNAL Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 FEATURES
 source 1..10244
 /organism="Thermotoga maritima"
 /db_xref="taxon:2336"
 gene 73..1104
 /gene="TM1744"
 CDS 73..1104

/gene="TM1744"
/note="similar to GB:U00096 SP:P75823 PID:1787095
PID:2225892 percent identity: 67.58; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="L-allo-threonine aldolase"
/protein_id="AAD36809.1"
/db_xref="GI:4982322"
/translation="MIDLRSPVTAKPTTEEMKKAQAEVGGDVYGEDPTINELRLAA
ETREKKAALVPVSGTMCNOVSIMAHNRGDVILLADSHIFMYEVGAAVAIVGYMPH
VPGKNGAMDPPDVKKAIRPRNTHFRPSLSIAIEMTHNRSGRVRVPLEMKICITAKE
HGINVHIDGARIFENASIASGVAVKEVIGADSVNFCISKGCAPAGSVYVDDRFIER
AKARKMIGGMRQAGVLAAGIATLRKMDRLKEDENARFLAKLKEIGSYNVED
VKTMVILRTDNLKVNAGFIETALRNSVILANAVSDIEIRLVTHKDVSRNDIEALNI
FEKLFKES"

1183. 1806
/gene="TM1745"
1183. 1806
/gene="TM1745"
/note="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36810.1"
/db_xref="GI:4982323"
/translation="MKRFVVMVLVTVALGVFAAGYRGFAERAPEROPMGVAQDFG
CGGIGVGNAPMYERQPMARERGVGLKIQIRDEQLKALKRIEALNNIDLD
DOLKEIVNAKETKEQLDLSKEKFEVQLHEEYVNALVRDRTAEDLKQIDDLRDEVO
DIKDFADKVDIITVDYKRFGRGKRLFLYLLIDEGEVLEEMQV"

2074. 4002
/gene="TM1746"
2074. 4002
/gene="TM1746"
/note="similar to GB:AE000783 percent identity: 56.54;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, periplasmic
oligopeptide-binding protein"
/protein_id="AAD36811.1"
/db_xref="GI:4982324"
/translation="MRKFVGFVLVLAFAFYSEVEVQLPMFGVTFNDFYVQAQHIL
AGCTFEELGEGVDPKSLGVKVPVAVCESYTSSEGLTFYFLRKDRMNGPEVT
LADFEFAKRAASPTPLPLMAKSPVOYIKKYNVYDCKGAKPLDELGVKRLDDYTLQVT
LSRPPSTINMLVUGAMPILHRTYOEHPEDMKRPYVPGNCPYIEESTPMKYEIVL
RNKYYVGPENVDRIVLKAGLGLQVYLAGIDVETIYAVGDIYFALNKQLSKELH
ESGVQWVGYETRSLSVFDDIRIRKALAMIDKRVLLDIVLGWALPTHAVCSFDS
ETAEAVKGIQYDEQAKRLAEGYVNGCPKPKVFYITGASDPVAEALVDQMKVYL
ITFEETENESGGSTLQMSDFVMAEPGCTIGPVNMQDSLSMQTADHTWFDYF
AEVRGKIYELKRESVLAKEGKTEEEKLEILNRAQELMEKMOEYVNSEPEFLKYY
FLPAPGELTFTEQINKLYEEMKATTDQKQLWRLAMRAYNOEYSVIYETGNMPT
NNEARRLMARILYSPYDKAVEYAKKLQQLCVDLATEVPLVYVKVYLAFLKGIYPI
KFSMGTFEFLFYLTVEK"

3999. 5024
/gene="TM1747"
3999. 5024
/gene="TM1747"
/note="similar to GB:I42023 SP:PA5054 PID:1006439
PID:1221243 PID:1205369 percent identity: 65.90;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36812.1"
/db_xref="GI:4982325"
/translation="MNSPEWDPFGLLNNSNGVDMRLFKLLKRLTLTAISMVAVI
ITYVLMVLAGNFEELQVADALARYTDPDPAYQATLKGFEERYGLNPLMLQAIY
LKGAVLFKRGSPSDPARNIEDLKEKPTTFLLASLILFALVGVPLGLIAALKN
TWIDYRAMVSVIGVAPSVYAVFLILFISYILGMLPTSGMEGIRTKIPLTIALAG
PLASVARTFVSLDITLNDQFLNATAKGGDDTVIKMHALRSMPIYITIVGPMAY
LMVGTWVENIFRIPGLGLFANAAYTRDIPLLVSTIFLALVIMNLIVDLVAIL

DPRIKLD"
5042. 5911
/gene="TM1748"
5042. 5911
/gene="TM1748"
/note="similar to GB:M57689 SP:P24139 GB:X56347 PID:143606
PID:40007 percent identity: 64.69; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36813.1"
/db_xref="GI:4982326"
/translation="MLRKNKTESGYKAKFWLRFKKKNKAIVGGVFLILALAILAP
YIAPYDPDEHYIAPFGPSKDFIRGDMALGRPLFSRIILYSRNCIIGEGSOPVLI
IGGLIYAVAFKPGKGIIDKFTMSIVDIDMEAFPEPLFVITVTLAAGRLFTIPLAIGTG
WAGMARLVQRQVLYLKNSEFVEAKAGASTFIIRKHLIPNMIGIYLVNLAFYPGA
MTESGLAVIGMGVRRPMPMGNLIGIGIMMAFPFLIFPAVTFATLISFTLAD
GLRDAFNPDSIS"
5908. 6882
/gene="TM1749"
5908. 6882
/gene="TM1749"
/note="similar to SP:P04285 GB:X05491 GB:X52093 PID:47805
percent identity: 69.65; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36814.1"
/db_xref="GI:4982327"
/translation="MVELLVNMLKVEFHRVEGIVKAVDGISYKLNKGSLGIVGSG
/SGKSVSILRLINRNGRIYDGEALFGLDKLKNKEELRNIRIGDIIIFONPMTS
LNPIIRIGVIMEPIIWHRLMKNEAEERAIETELLERGIPEKRLNFPQSGGAR
QRVITAMALACHKRLIADDEFTTALDVTIQAOIMEILOETKEEYGSVFIFTHDSLVA
TNFCDRLITTAGKIVEAEPEELKTLPHFTYRGLNLTLEIGSGKRLVPIPGNP
NPKRPSGCKPHRCSFAMEICQREPEPLVNISBNHVAACHLNGESK"

6852. 7838
/gene="TM1750"
6852. 7838
/gene="TM1750"
/note="similar to SP:PA2065 PID:677944 GB:AL009126 percent
identity: 76.43; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36815.1"
/db_xref="GI:4982328"
/translation="MPEVQGIKMKPLQTVYDLKKPYPOCKRILKANDGISIEIKKE
TLGLVSGCGCKSTLGRWILKILPLPDGKLIFFEEKDITNLDMKMKYRKMKQITRD
PLGSLNPQWYGRITIEDPLLIHKIGTKERKRVDELDMVIGIGREFINSFHPESG
QOORIGIALALANLKPFTVCEPVSALDVSIOADIIDLEPIQCKMISYLFIAHNL
VYEHSHKVAWYKCIYEVGDVADKIFLNPHTRALIKVPIKIPDGGOKORFYSLK
GELSPIDLPKRCRQTRCTKKAICFEKEPELLEVENHVSCHLVRSG"

7898. 8851
/gene="TM1751"
7898. 8851
/gene="TM1751"
/note="similar to SP:P23340 percent identity: 52.35;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endoglucanase"
/protein_id="AAD36816.1"
/db_xref="GI:4982329"
/translation="AGVDPFENKLTIGKINIGNALPANEQDMGVYKIDFPIIKE
/AFSHVRIETKSHTAIVAPPKINDREFKRYDEVINALRGLAVVNIHHYELMN
DKDEHKERELAMKQIADRYKYDEPTLFEILINEHGLNTEPKNELEELKYLIRSI
DKEHIIIGTAWGGLIALEKISVPRKEKNSIVTTHINPEFTHOGAEWEGSEKWL
GRKMSPPDOKHLIEEFNIEBWSKKNRPYIYIGFAYRRADLESRIKMTSVFREM

gene
EKRRSWAYWFCGSGFYVDTLRKTNKDLLEALIGDSIE"
8844..9833
/gene="TM1752"
8844..9833
/gene="TM1752"
/note="similar to SP:P23340 percent identity: 51.51;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endoglucanase"
/protein_id="AAD36817.1"
/db_xref="GI:4982330"
/translation="NNNTTPRWGRFNLLEASIKSTGNFKEEDFLMAQWDFNVRIP
MCHLMSDRGNFPIIREDFEKIDRIVTGEKYGIIHICISLRAGSYNKEVEKTN
LWKDETQAEAFIHMNSFIARRYKGISSTHLENLINEPEPDQJMSVEDHNSLIKRT
ITEIRKIDPERLIIDIGYGNIPVDLTIENYVSCRGYIFSVTHYKAEVDSKDF

Query Match 99.8%; Score 952.4; DB 1; Length 10244;
Best Local Similarity 99.9%; Pred. No. 5.5e-203;
Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGCTGTTGATCCCTTTGAAGACAAATTTGGAGAGCGATTAATATAGGAAT 60
|||||
Db 7898 ATGGGCTGTTGATCCCTTTGAAGACAAATTTGGAGAGCGATTAATATAGGAAT 7957
OY 61 GCGCTTGAAGACCAATGAGGAGAGCTGGGAGTGTGATTAAGATGATCTTCGAC 120
|||||
Db 7958 GCGCTTGAAGACCAATGAGGAGAGCTGGGAGTGTGATTAAGATGATCTTCGAC 8017
OY 121 ATTATTAAGAACCGCGTTCTCTCATGTTGCAATTCACATTAAGATGAGTACGCGCT 180
|||||
Db 8018 ATTATTAAGAACCGCGTTCTCTCATGTTGCAATTCACATTAAGATGAGTACGCGCT 8077
OY 181 TAGCGCTTCCCTCTATTAATCATGATCGCTCTCAAAAAGATGATGATGAAGGATA 240
|||||
Db 8078 TAGCGCTTCCCTCTATTAATCATGATCGCTCTCAAAAAGATGATGATGAAGGATA 8137
OY 241 AAGGAGCCCTGAAAAGAGAGTGGCTGTGCTAATTAATTCATCACTACGAGAGATTA 300
|||||
Db 8138 AAGGAGCCCTGAAAAGAGAGTGGCTGTGCTAATTAATTCATCACTACGAGAGATTA 8197
OY 301 ATGAATGATCCAGAGACACACAAAGAAATTTCTGCTCTTTGGAAACAAATGCTGAT 360
|||||
Db 8198 ATGAATGATCCAGAGACACACAAAGAAATTTCTGCTCTTTGGAAACAAATGCTGAT 8257
OY 361 CGTTATTAAGACTATCCGAAACTATTTTGAATCTGATGAACCTCAGCGAAT 420
|||||
Db 8258 CGTTATTAAGACTATCCGAAACTATTTTGAATCTGATGAACCTCAGCGAAT 8317
OY 421 CTTACTCCGAAAAAATGATGAATGAATCTGTAAGAGCTCTAAAGTTATAGATCAAT 480
|||||
Db 8318 CTTACTCCGAAAAAATGATGAATGAATCTGTAAGAGCTCTAAAGTTATAGATCAAT 8377
OY 481 GACAAAAACACACTATTAATATAGCACAGCTGAATGGGGGGTATATGCTCTTGA 540
|||||
Db 8378 GACAAAAACACACTATTAATATAGCACAGCTGAATGGGGGGTATATGCTCTTGA 8437
OY 541 AAGCTGCTGTCCCAAAATGGGAAAAAATTCATAGTTACAATTCACACTACATCCCT 600
|||||
Db 8438 AAGCTGCTGTCCCAAAATGGGAAAAAATTCATAGTTACAATTCACACTACATCCCT 8497
OY 601 TTGGAATTTACCATCAGAGAGCTGAGTGGAGAGATCTGAGAAATGTTGGGAGAA 660
|||||
Db 8498 TTGGAATTTACCATCAGAGAGCTGAGTGGAGAGATCTGAGAAATGTTGGGAGAA 8557
OY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTATAGAGAA 720
|||||
Db 8558 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTATAGAGAA 8617
OY 721 TGGTCAAAAAAGACCAATTTTACATAGGTGAGTTGGTCCCTACGAAAAAGCT 780
|||||
Db 8618 TGGTCAAAAAAGACCAATTTTACATAGGTGAGTTGGTCCCTACGAAAAAGCT 8677

OY 781 GACCTTGAATCAAGATTAATATGACCTCTTGTGCTTCGCGAAATGAGAAAGAGAG 840
|||||
Db 8678 GACCTTGAATCAAGATTAATATGACCTCTTGTGCTTCGCGAAATGAGAAAGAGAG 8737
OY 841 TGGAGCTGGGCAATCTGCGAATTTTGTCCGGTTTGGTGTATATCTGACAGAAA 900
|||||
Db 8738 TGGAGCTGGGCAATCTGCGAATTTTGTCCGGTTTGGTGTATATCTGACAGAAA 8797
OY 901 ACTGGAATTAAGATCTTTAGAGCTTTAATATGAGAGAGATGACATTTGATA 954
|||||
Db 8798 ACTGGAATTAAGATCTTTAGAGCTTTAATATGAGAGAGATGACATTTGATA 8851

RESULT 5
AR205090 1083 bp DNA linear PAT 20-JUN-2002
LOCUS
DEFINITION Sequence 13 from patent US 6368844.
ACCESSION AR205090
VERSION AR205090.1 GI:21502584
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1083)
AUTHORS Bylina, E.J.
TITLE Glycosidase enzymes
JOURNAL Patent: US 6368844-A 13 09-APR-2002;
FEATURES Location/Qualifiers
source
1..1083
/organism="unknown"
BASE COUNT 367 a 189 c 242 g 285 t
ORIGIN

Query Match 24.7%; Score 235.8; DB 6; Length 1083;
Best Local Similarity 54.2%; Pred. No. 1.4e-42;
Matches 502; Conservative 0; Mismatches 422; Indels 3; Gaps 1;

OY 14 CTTTGAAGAGAACAAATATTGGAGAGCGCATTAATATAGAAATGCGCTTGAAGCAC 73
|||||
Db 149 CATTGAATTAACAAATATGATGATGAAGAGATTAATATGGAATGCTTTGAAGCTC 208
OY 74 CAATAGGAGAGACTGGGAGTGTGATTAAGATGATGATGATGATGATGATGATGAT 133
|||||
Db 209 CTTTCAAGAGCGCTGGGAGTGTGATTAAGATGATGATGATGATGATGATGATGAT 268
OY 134 CCGGTTCTCTCATGTTGCAATTCATTAAGATGATGATGATGATGATGATGATGATGAT 193
|||||
Db 269 GGGGATTGATCTGTAGGATTCCTCATAGATGTCGACATATATATCCGAAAAAGCCAC 328
OY 194 CTTATTAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 253
|||||
Db 329 CATATATATTTGACAGAGATTTCTCGAAAGATTAACCATGTTGTCGATAGGGCTCTTG 388
OY 254 AAGAGAGACTGGCTGTCTGATTAATATTTCTGATCACTACGAGAGATTAATGATGATCAG 313
|||||
Db 389 AGAATTAATTAACAGTAATCATCATACGACCATTTTGAAGACTCTATCAAGAACCGG 448
OY 314 AAGAACACAAAGAAATTTCTGCTTTGGAACAAATGCTGATTCGTTATAAGAGACT 373
|||||
Db 449 ATTAATACGCGCATGTTTGGTGAATTTGGAGACAGATGGAATTTCTTTAAAGATT 508
OY 374 ATCCCAAACTCATTTTGTAAATCTGAATGACCTCAAGGAATCTTACTCCGAGAA 433
|||||
Db 509 ACCCGAAATATCTGTTCTTTGAAATCTACACAGAGCTGCTCGAACTTGACAGCTGAAA 568
OY 434 AATGAATGACTGCTTGAAGAGCTCTAAAGATTAATGATCAATTTGAACAAAGACACA 493
|||||
Db 569 AATGAATGACTGCTTGAAGAGCTCTAAAGATTAATGATCAATTTGAACAAAGACACA 628
OY 494 CTAATATTTAGGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 553
|||||
Db 629 TTGCTATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 688

-10_signal
CDS
897..902
/gene="cele"
1027..3471
/gene="cele"
/codon_start=1
/transl_table=1
/product="endo-beta-1,4-glucanase"
/protein_id="AA03224.1"
/db_xref="GI:144770"
/translation="MKRIIVLCVLMVLSISGFSVAAPVKFGVSGTKLADAG
NELVWGRDLSAIDLYKEIKITGNLNTDAPETETAGNRTKAMIEKREMGFA
VRVPTWDTHTGPAVDKIDEMLNREVEVNYVLDGMAITLHNDNMITVYN
EORSEKLVKWEQIATRFKDYDHLLETNNEPEVSEPMAGCTENRDNREN
LAVNTIRASGNDKRIILVPTNAATGLDALNDLVI PNDSRVISVIAHSEFFA
MDVGTSGWSDYKASLSELDIYNRFVNGRAVILIGEGTIDKNLSRVAHAH
YARAVSGIIVEMWMDNGYNGDAETALNRLTISWYPEIIVQALMGAGVEPLVS
PTPTPTPSPPTVNTAIIYGVNDGKINSTCTMKNRYILGIEEFPSPGIIA
DVNDLKNSTDIYLMKKYILRSIDGPNPDESDPDESDPDESDPDESDPDESDP
AMSGSNELNFTGTEASVYIKSGENFQALVDNCPIDPESVNTTSVKLVGLAES
AAHLVLMKRTASLGEVDFGSGKLAPKPLEKIEFTIGDSITTCAGNCTSK
EOSTPKNENSVMSYAAITARNLASAMMIAMSGIGLTMNGAGAPLIMRYPTTL
YSGVMDFSKYVQVYVNIIGTNDSTFADKTKVAYKNLISEVRNRYDAHIFCC
VGPLMGTGLDCRSYVEVNDNCRSGDLKYFEVEFPODSTGYGDMHPSIATHQ
LMARLTAIRKIKGM"

sig_peptide
1027..1128
/gene="cele"
mat_peptide
1129..3468
/gene="cele"
/product="endo-beta-1,4-glucanase"

BASE COUNT 1074 a 632 c 811 g 954 t
ORIGIN

Query Match 8.0%; Score 76; DB 1; Length 3471;
Best Local Similarity 51.1%; Pred. No. 6.8e-07;
Matches 240; Conservative 0; Mismatches 215; Indels 15; Gaps 2;

28 AAAATATTGGAAGAGCCTTATATAGAAATGCCCTGAAACACCAATGAGGAGAC 87
1210 AAAGAAATTAATAATCGATGGAATTTGGAAATACCTTGATCGCTCCACAGACTGCC 1269

88 TGGGGA---GTGATATAAAGATGACTTCTTGACATTTTAAAGAGCGGCTTCTCT 144
1270 TGGGGAATTCAGAGACACCAAGGAGATGATAGAAAGGTAAGGAGGCTTTAAT 1329

145 CATGTTCGATTCCTCAATAGATGAGTACGACGCTTACGCTTCTCTCTTAAATC 204
1330 GCCGTGAGTGGCTCTTGCTTACCTGGATACGACATCGACCTGCTCGGACTTAAAT 1389

205 ATGATCGCTTCTTCAAAAGAGTGTGATGATGAAAGGAGCGCTGAAAGAGACTG 264
1390 GACGAGATGCTGGAACAGAGTGTGAGAGTGTGAAACTATGCTTGAAGCTGATG 1449

265 GCGTTCCTTAAATATTCATCACTACGAGAGTAAATGATGATCCAGAGAACAGAG 324
1450 TACGCGATCAATTAATCTTCCATGACATACATGATTTACTTACATATGCCAATGAG 1509

325 -----GAAAGATTCTTCTCTTGGAAACAAATGCTGATCGTTATTAAGAC 372
1510 CAAGAGTAAAGAAACTTGAAGTTTGGGAAACAAATAGCAACCCGTTTAAAGAT 1569

373 TATCCGCAAACTATTTTGAATCTGAAATGACCTCAGCAAAATCTTACCCGGA 432
1570 TATGACGACCATTTTGTGTGTGAGACATGAAACCGGAGAGAGTAGTTCACCTATG 1629

433 AAATGATGAACTGCTTGAAGAGCTCTAAAGTATTAAGATCAATTGA 482
1630 GAATGATGAGCGGAGAGCTATGAAACCGAGATGTGATTAACAGATTAA 1679

RESULT 11
AF015249 1451 bp mRNA linear PLN 23-MAR-2000
LOCUS AF015249
DEFINITION Orpinomyces joyonii cellulase (celB2) mRNA, complete cds.

ACCESSION AF015249
VERSION AF015249.1 GI:2353006
KEYWORDS
SOURCE
ORGANISM Orpinomyces joyonii.
REFERENCE
AUTHORS Orpinomyces joyonii
Eukaryota: Fungi: Chytridiomycota; Neocallimastixales;
Neocallimastixaceae: Orpinomyces.
1 (bases 1 to 1451)
AUTHORS Qiu, X., Sellinger, B., Yanke, L., and Cheng, K.
TITLE Isolation and analysis of two cellulase cDNAs from Orpinomyces
joyonii
JOURNAL Gene 245 (1), 119-126 (2000)
MEDLINE 20179691
PUBMED 10713452
REFERENCE
AUTHORS Qiu, X., Sellinger, L.B., Yanke, L.J. and Cheng, K.-J.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-1997) Lethbridge Research Centre, Agriculture
Canada, P.O. Box 3000, Lethbridge, AB T1J 4B1, Canada
Location/Qualifiers
1. 1451
/organism="Orpinomyces joyonii"
/strain="SC4"
/db_xref="taxon:48250"
1. 1451
/gene="celB2"
265..1431
/gene="celB2"
/codon_start=1
/product="cellulase"
/protein_id="AA069348.1"
/db_xref="GI:2353007"
/translation="MKFGMLNLTLDNOCITGLINSKDOTASETCGNGKPKTDEMFV
MIDNQVFLPIPTWGSCHFADEDEGHKIDEMKRVREVDIAINCAEPIYLIHETN
NHASETLEAKVILADINNOIAEERDVVERILFEGLENEPKRNDIPVWNGDEGV
DAVAMEVEFLKTIIRASGNNPKRHLIIPYAAATDENSEFKRYPGDDKLIVSHN
YAPYFALNNGDGAVEREDPAOGKDLDMSSILIKRFTPDGIPVIGEEYGAARONTE
ERAKWAEYKVTALIGVPOVMDNGIFEGTGRFLDRKRLIYPSIVALALGR
GLEVEYKHAENAVESGMAETLIEPCCSEBENTRVATDENKMGVNGWGCIP"

BASE COUNT 456 a 221 c 337 g 437 t
ORIGIN

Query Match 7.9%; Score 75.4; DB 8; Length 1451;
Best Local Similarity 50.6%; Pred. No. 1e-06;
Matches 215; Conservative 0; Mismatches 201; Indels 9; Gaps 1;

99 GATTAAGATGAGTCTTCGACATTTATAAAGAACCCGCTTCTCTCATGTGCAATTC 158
372 GACTACTGAAGATGATGTCACAGATTATGATGATCAATTCATGTTCCGATTC 431

159 AATAAGATGAGTACGACGCGCTTACGCGCTTCTCTTATTAATCATGATGCTTCT 218
432 AACTACTGCTGCTGGTCACTTTGCTGATGATGATGATGATGATGATGATGATGAT 491

219 CAAGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 278
492 GAAAGAGTTCACAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 551

279 TATTTCATCA-----CTAGAGAGTATTAATGATGATGATGATGATGATGATGAT 329
552 CATTCATCAAGAAACCTGAGAACCTGATGATGATGATGATGATGATGATGATGAT 611

330 ATTCTGCTCTTGGAAACAAATGCTGATGATGATGATGATGATGATGATGATGAT 389
612 TTTAGCTGATATTTGGAATCAATTTCTGAAAGATCAAGATGATGATGATGATGAT 671

390 TTTTGAATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 449
672 TTTCGAGAGTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 731

450 TGAGAGAGCTTAAAGTATTAAGATGATGATGATGATGATGATGATGATGATGAT 509
732 TAAAGAGGTTGGATGCTGTTATGCTATGATGATGATGATGATGATGATGATGAT 791

Db	398	TTCAATGAAGTGTGCATTATGCTCTTAACTGGTGGTTATGCCATCTTAAACAATTCACC	457
Qy	287	ACTACG-----AGGAGTTAATGAATGATCCAGAGAACACAGGAAAGATTCTTG	337
Db	458	ATGAACCTTGGAAATATGCTCTTCCAAAAGAAATTTAGAGAGTGCCAAAAAGATCTTAGTTG	517
Qy	338	CTCTTTGGAAACAATTTGCTGATTCCTTAAAGATCTCCGAAACCTGATTTTGTGAA	397
Db	518	CCATCTGGAAACAATTCCTGCTGAATTTGCTGATTTATGATGAACACATTTATTTTGCAG	577
Qy	398	TTCTGAATGAACCTACGGAATCTTACTCCGGAATAATGGAATGAACCTGCTTGAGGAG	457
Db	578	GAATGAATGAACCAAGAAAGGTTGGGGATCCAGCTGAATGGACAGGTGATCAAGAAG	637
Qy	458	CTCTAAAGATTAAAGATCAATTGACACAAAAGACACATTAATTTATAGGCACAGGTGAAT	517
Db	638	GTTGGAAATTCCTCAATGAATAATGAATGCCCTTTTGTTAAACATATCTGTGCCACTGGAG	697
Qy	518	G	518
Db	698	G	698

[illegible]

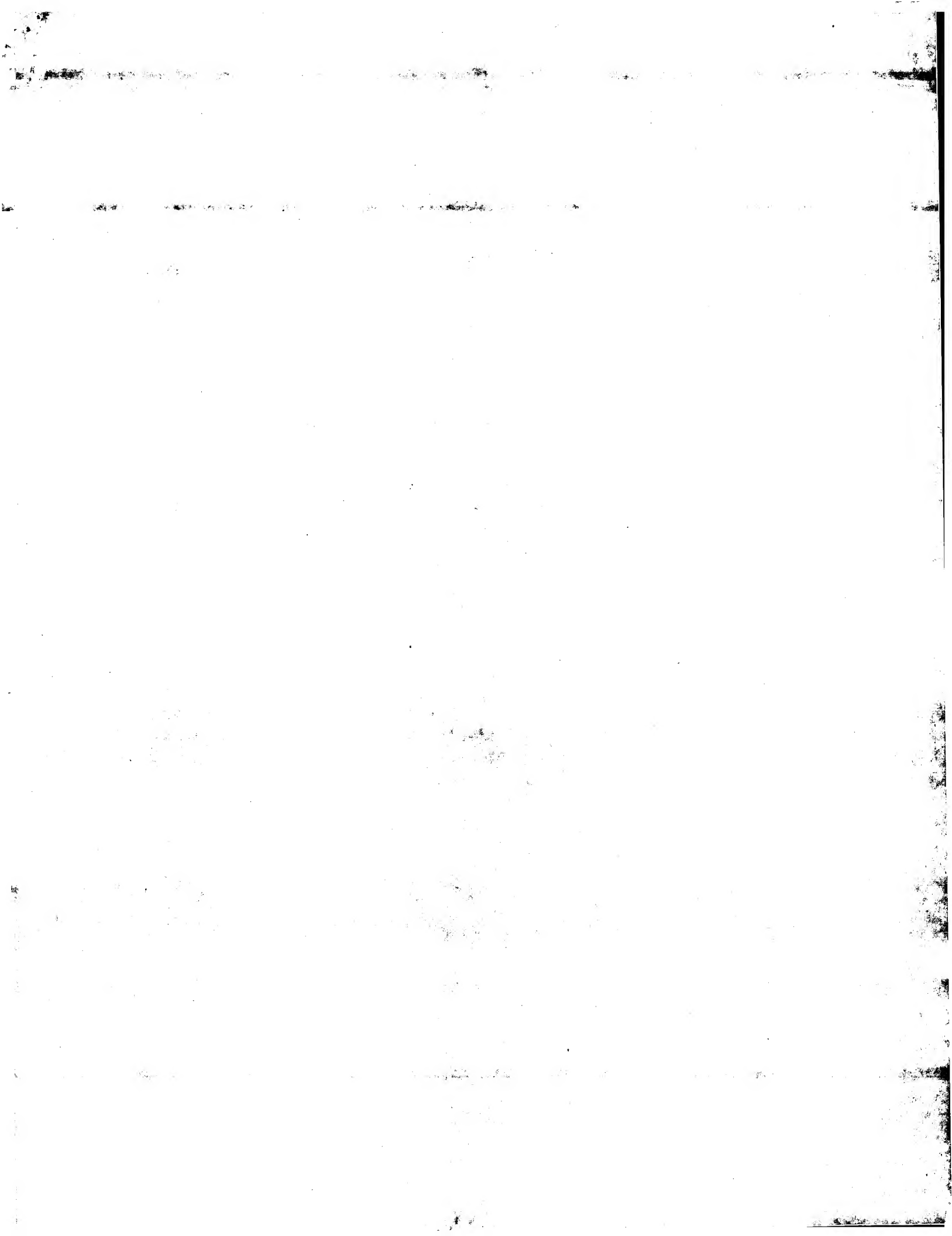
Db 637 GTGGCAATTTGTCATCAATGAATGACCTTTTCGTTAAACTATTCGTGCACCTGGAG 6366

QY 518 G 518

Db 697 G 697

RESULT	15
LOCUS	P0Q277483
DEFINITION	P0Q277483 5502 bp mRNA linear PLN 29-AUG-2006
ACCESSION	AJ2277483
VERSION	AJ2277483.1 GI:8052315
KEYWORDS	cel5A gene; endoglucanase.
SOURCE	Pyromyces equi.
ORGANISM	Pyromyces equi Eukaryota; Fungi; Chytridiomycota; Neocallimastixales; Neocallimastixaceae; Pyromyces.
REFERENCE	1 (bases 1 to 5502)
AUTHORS	Eberhardt,R.Y., Gilbert,H.J. and Hazlewood,G.P.
TITLE	Primary sequence and enzymic properties of two modular endoglucanases, Cel5A and Cel5A', from the anaerobic fungus

[illegible]



GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 07:29:37 ; Search time 191 Seconds
(without alignments)
11248.201 Million cell updates/sec

Title: US-09-880-729A-1

Perfect score: 954
Sequence: 1 atgggtgtgtagcttcttga.....gaggagatagcatgtaataa 954

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	954	100.0	954	20	AAZ21563
2	954	100.0	954	20	AAZ21563
3	954	100.0	954	21	AAZ46759
4	954	100.0	954	22	AAZ46759
5	235.8	24.7	1083	19	AAV36199
6	232.6	24.4	1083	18	AAV36199
7	73	7.7	1826	19	AAV29477
8	73	7.7	1826	19	AAV37413
9	68.6	7.2	1532	21	AAA64503

10	63.2	6.6	4098	15	AAQ55036
11	62.8	6.6	1170	22	AAE83293
12	62.8	6.6	1182	22	AAE83296
13	62.8	6.6	2110	22	AAE83292
14	49.4	5.2	1614	20	AAZ05950
15	48.2	5.1	2777	17	AAZ41849
16	42.6	4.5	32184	22	ABA20589
17	42.6	4.5	32184	22	ABA20590
18	42.6	4.5	38372	22	AAK84952
19	42.6	4.5	38348	22	AAK84953
20	42.4	4.4	6467	24	ABN80169
21	42	4.4	2172	20	AAZ61594
22	42	4.4	2262	20	AAZ61593
23	41.6	4.4	3740	15	AAQ54682
24	40.8	4.3	6290	24	ABL33046
25	40.6	4.3	7442	22	AAZ46686
26	40.2	4.2	585	24	ABQ60399
27	40.2	4.2	7317	22	AAZ45342
28	40.2	4.2	7317	24	ABK28173
29	40.2	4.2	16228	24	ABL70459
30	40.2	4.2	16228	24	AAZ61424
31	39.8	4.2	6392	24	ABL32685
32	39.8	4.2	6392	24	ABL34507
33	39.2	4.1	2166	11	AAQ06781
34	38.8	4.1	6853	24	ABL54331
35	38.8	4.1	6853	24	ABL32310
36	38.2	4.0	5501	24	ABL32054
37	38	4.0	2623	23	ABL12172
38	38	4.0	4288	23	ABL05576
39	38	4.0	8067	24	ABN80224
40	38	4.0	580073	18	AAZ58840
41	37.8	4.0	1092	22	AAZ16819
42	37.8	4.0	1284	23	AAZ70787
43	37.8	4.0	6740	24	ABL33147
44	37.6	3.9	611	23	ABV51437
45	37.6	3.9	6863	22	AAZ46618

ALIGNMENTS

RESULT 1	AAZ21563	standard; cDNA; 954 BP.
ID	AAZ21563	
AC	AAZ21563	
DT	01-DEC-1999	(first entry)
DE	Carboxymethyl cellulase (CMCase) gene.	
KW	Carboxymethyl cellulase; CMCase; thermostable enzyme; hydrolysis;	
KW	beta 1 4 glycosidic bond; cellulose; plant biomass conversion; fuel;	
KW	chemical; detergent; textile industry; animal feed; waste treatment;	
KW	fruit juice extraction; clarification; brewing industry; ss.	
OS	Thermotoga maritima.	
FH	Key	Location/Qualifiers
FT	CDS	1..954
FT		/*tag= a
FT		/product= CMCase
FT		/note="Carboxymethyl cellulase"
XX	US5962258-A.	
XX	05-OCT-1999.	
XX	23-AUG-1995;	95US-0518615.
XX	23-AUG-1995;	95US-0518615.
XX	23-AUG-1995;	95US-0518615.
XX	(DIVE-) DIVERSA CORP.	

Sequence of celd c
P. rhizinflata egl
P. rhizinflata egl
P. rhizinflata egl
P. rhizinflata egl
Bacillus agardhae
Bacillus agardhae
Human nervous syst
Human nervous syst
Human nervous/haema
Human immune/haema
Human immune/haema
B. burgdorferi ant
B. burgdorferi ant
Potato sucrose pho
Human immune syste
Tumour suppressor
Human colon cancer
Chemically pretrea
DNA transcription
Chemically treated
Human gene regulat
Human immune syste
Human metastasis a
Sequence encoding
Chemically treated
Human immune syste
Human immune syste
Drosophila melanog
Drosophila melanog
Human chemically m
Mycoplasma genital
Paenibacillus pabu
DNA encoding novel
Human immune syste
Human prostate exp
Tumour suppressor

XX Mathur EJ, Lam DE;
 XX WPI: 1999-571269/48.
 DR P-PSDB: AAY39338.
 PT Recombinant polynucleotide sequence encoding a carboxymethyl cellulase
 useful for the conversion of plant biomass -
 XX
 PS Claim 11; Fig 1; 14pp; English.
 CC This sequence is a carboxymethyl cellulase gene (CMCase) is a
 CC thermostable enzyme (optimum temperature of 60 + degrees C) that
 CC catalyses the hydrolysis of the beta 1,4 glycosidic bonds in cellulose.
 CC The enzyme degrades cellulose by the hydrolysis of this bond. The CMCase
 CC enzyme has a molecular weight of approximately 35KD. CMCase is used to
 CC catalyse enzymatic degradation of cellulose and can be used for the
 CC conversion of plant biomass into fuels and chemicals. CMCase can also be
 CC used in detergents, the textile industry, animal feed, waste treatment
 CC and for the clarification and extraction of juices in the fruit juice and
 CC brewing industries. The CMCase gene and 15 bp fragments of it can be used
 CC as probes to isolate nucleic acid molecules encoding CMCase enzymes from
 CC genomic libraries or to screen for related sequences. Cells expressing
 CC the CMCase polynucleotide can be used as immunogens to produce antibodies
 CC which can then be used to isolate the enzyme from cells which express
 CC CMCase. The antibodies can also be used to screen for similar enzymes
 CC from other organisms and samples.

Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other;

Query Match 100.0%; Score 954; DB 20; Length 954;

Best Local Similarity 100.0%; Pred. No. 1.6e-244;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTGTTGATCTTTTGAAGAACAAATATTGGGAGAGCGCTTATATAGCAAT 60
 DB 1 ATGGGTGTTGATCTTTTGAAGAACAAATATTGGGAGAGCGCTTATATAGCAAT 60
 QY 61 GCGCTTGAAGCAACCAATAGGAGAGCTGGGGAGTGTATATAAATGATGCTTCTCGAC 120
 DB 61 GCGCTTGAAGCAACCAATAGGAGAGCTGGGGAGTGTATATAAATGATGCTTCTCGAC 120
 QY 121 ATTATTAAGAGAGCGGCTTCTCATGTTTCAATTCATTAAGATGAGATGACGCT 180
 DB 121 ATTATTAAGAGAGCGGCTTCTCATGTTTCAATTCATTAAGATGAGATGACGCT 180
 QY 181 TACGCGTTTCTCTTATTAATATCATGATCGCTTCTTCAAAAGATGATGATGATGAT 240
 DB 181 TACGCGTTTCTCTTATTAATATCATGATCGCTTCTTCAAAAGATGATGATGATGATGAT 240
 QY 241 AACGGAGCCCTGAAAAGAGAGACTGGCTGTATATAATTCATCATCATCGAGGATTA 300
 DB 241 AACGGAGCCCTGAAAAGAGAGACTGGCTGTATATAATTCATCATCATCGAGGATTA 300
 QY 301 ATGAATGATCCAGAGAACCAAGAAAGATTTCTTGGAAACAAATTTGCTGAT 360
 DB 301 ATGAATGATCCAGAGAACCAAGAAAGATTTCTTGGAAACAAATTTGCTGAT 360
 QY 361 CGTTATTAAGACTATCCCGAACTATTTTGAATTTGAAATTCGATGAACCTCAGCAAT 420
 DB 361 CGTTATTAAGACTATCCCGAACTATTTTGAATTTGAAATTCGATGAACCTCAGCAAT 420
 QY 421 CTGACTCCGGAATAATGATGAACTGCTGAGGAAGCTCTAAAGTTATTAAGTCAATT 480
 DB 421 CTGACTCCGGAATAATGATGAACTGCTGAGGAAGCTCTAAAGTTATTAAGTCAATT 480
 QY 481 GACAAAAAGCACACTATATATATAGCAACGCTAAATGGGGGATATATCGCCCTGAA 540
 DB 481 GACAAAAAGCACACTATATATATAGCAACGCTAAATGGGGGATATATCGCCCTGAA 540
 QY 541 AAACGTGCTGCCCAAAATGGGAAAATTCATTAATTAATCAATCACTACTACTACTCT 600
 DB 541 AAACGTGCTGCCCAAAATGGGAAAATTCATTAATTAATCAATCACTACTACTACTCT 600

QY 601 TTGCAATTTACCCATCAAGAGCGTGAAGGATCTGAGAAATGTTGGAGA 660
 DB 601 TTGCAATTTACCCATCAAGAGCGTGAAGGATCTGAGAAATGTTGGAGA 660
 QY 661 AAGTGGGATCTCCAGATGATCAAGAAACATTTGATAGCAATTTATAGAGA 720
 DB 661 AAGTGGGATCTCCAGATGATCAAGAAACATTTGATAGCAATTTATAGAGA 720
 QY 721 TGGTCAAAAAGAACCAAGACCAATTTACATAGCTGATGCTGACAGAAAGCT 780
 DB 721 TGGTCAAAAAGAACCAAGACCAATTTACATAGCTGATGCTGACAGAAAGCT 780
 QY 781 GACCTGATCAATCAATTAATGACCTCTTGTCTGCGCAAAATGAGAAAGAGA 840
 DB 781 GACCTGATCAATCAATTAATGACCTCTTGTCTGCGCAAAATGAGAAAGAGA 840
 QY 841 TGGAGCTGGGATCTGGAATTTTGTCCGCTTTGCTTTATATGATCTGACAAA 900
 DB 841 TGGAGCTGGGATCTGGAATTTTGTCCGCTTTGCTTTATATGATCTGACAAA 900
 QY 901 ACCGTGAATTAAGATCTTTTGAAGCTTTATATAGAGATGATGATTAATA 954
 DB 901 ACCGTGAATTAAGATCTTTTGAAGCTTTATATAGAGATGATGATTAATA 954

RESULT 2

AAK85879
 ID AAK85879 standard; cDNA; 954 BP.

AAK85879;

10-SEP-1999 (first entry)

DE cDNA encoding a carboxymethyl cellulase enzyme.

XX Carboxymethyl cellulase enzyme; hydrolysis; cellulose; plant biomass;

KW fuel; chemical; detergent; textile industry; animal feed;

KW waste treatment; fruit juice; brewing industry; clarification;

KW juice extraction; ss.

OS Thermotoga maritima.

PN US5925749-A.

PD 20-JUL-1999.

XX 23-AUG-1995; 95US-0518615.

XX 23-AUG-1995; 95US-0518615.

XX 24-APR-1996; 98US-0066075.

PA (DIVE-) DIVERSA CORP.

PI Lam DE, Mathur EJ;

XX WPI: 1999-429058/36.

DR P-PSDB: AAY23764.

XX New carboxymethyl cellulase enzyme is useful in cellulose digestion

PT Claim 1; Fig 1A-C; 13pp; English.

XX The present sequence encodes a carboxymethyl cellulase enzyme from

CC Thermotoga maritima. Oligonucleotide probes derived from the

CC polynucleotide sequence can be used to screen a library of genomic

CC DNA for recovery of the polynucleotide, as PCR primers to synthesize

CC full length polynucleotides or as diagnostic probes useful for

CC screening for similar enzymes from other organisms and samples.

CC The carboxymethyl cellulase enzymes can be used to catalyze hydrolysis

CC of cellulose for the conversion of plant biomass into fuels and

CC chemicals, for use in detergents, the textile industry, in animal feed,

CC waste treatment and in the fruit juice/brewing industry for the

CC clarification and extraction of juices.

XX Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other;

Query Match 100.0%; Score 954; DB 20; Length 954;
 Best Local Similarity 100.0%; Pred. No. 1.6e-244;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ATGGGTTGATCCTTTTGAAGCAAAATTTGGGAAGGCACTTAATATAGCAAAAT 60
   |||||||
DB 1 ATGGGTTGATCCTTTTGAAGCAAAATTTGGGAAGGCACTTAATATAGCAAAAT 60
OY 61 GCGCTTGAAGCAACCAATGAGGAGACTGGGAGTGTATAAAGATGAGTCTTGAC 120
   |||||||
DB 61 GCGCTTGAAGCAACCAATGAGGAGACTGGGAGTGTATAAAGATGAGTCTTGAC 120
OY 121 ATTATTAAGAAGCCGGTTTCTCATGTTCCGAATTCCAATAGATGAGTACGACGCT 180
   |||||||
DB 121 ATTATTAAGAAGCCGGTTTCTCATGTTCCGAATTCCAATAGATGAGTACGACGCT 180
OY 181 TACGCTTTCCTCTTATTAATCATGATCGCTTCTCAAAAGAGTGGATGAAGTATA 240
   |||||||
DB 181 TACGCTTTCCTCTTATTAATCATGATCGCTTCTCAAAAGAGTGGATGAAGTATA 240
OY 241 AACGAGCCCTGAAAAGAGACTGGCTGTGCTATAAATATTCATCAGTACGAGAGTTA 300
   |||||||
DB 241 AACGAGCCCTGAAAAGAGACTGGCTGTGCTATAAATATTCATCAGTACGAGAGTTA 300
OY 301 ATGATATGATCCAGAAACACAAAGAAAGATTTCTGCTTTGGAAACAAATTCGTGAT 360
   |||||||
DB 301 ATGATATGATCCAGAAACACAAAGAAAGATTTCTGCTTTGGAAACAAATTCGTGAT 360
OY 361 CATTATAAGACATCCCGAACTATTTTGAATTTCTGAATGACCTCAGCAAAAT 420
   |||||||
DB 361 CATTATAAGACATCCCGAACTATTTTGAATTTCTGAATGACCTCAGCAAAAT 420
OY 421 CTCTACTCGGAAAATGGAATGAGTCTGAGAAAGCTCTAAAGTTATTAAGTCAAT 480
   |||||||
DB 421 CTCTACTCGGAAAATGGAATGAGTCTGAGAAAGCTCTAAAGTTATTAAGTCAAT 480
OY 481 GACAAAACACACTATTAATATAGGCACAGCTGAATGGGGGTATATCTGCCCTTGA 540
   |||||||
DB 481 GACAAAACACACTATTAATATAGGCACAGCTGAATGGGGGTATATCTGCCCTTGA 540
OY 541 AAACGTCTGTCCCAAAATGAGAAAATTTCTATAGTTACAATTCACACTCAATCC 600
   |||||||
DB 541 AAACGTCTGTCCCAAAATGAGAAAATTTCTATAGTTACAATTCACACTCAATCC 600
OY 601 TTGCAATTTACCATCAAGAGCTGAGTGGGTGGAAGATCTGAGAAATGGTTGGAAGA 660
   |||||||
DB 601 TTGCAATTTACCATCAAGAGCTGAGTGGGTGGAAGATCTGAGAAATGGTTGGAAGA 660
OY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTATAGAAAGA 720
   |||||||
DB 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTATAGAAAGA 720
OY 721 TGGTCAAAAAGAACCAATTTTACATAGTGTGAGTTGGTCCCTACAGAAAAGCT 780
   |||||||
DB 721 TGGTCAAAAAGAACCAATTTTACATAGTGTGAGTTGGTCCCTACAGAAAAGCT 780
OY 781 GACCTTGAATCAAGATTAATATGAGCTCTTGTCTCGGAAATGAGAAAAGAGA 840
   |||||||
DB 781 GACCTTGAATCAAGATTAATATGAGCTCTTGTCTCGGAAATGAGAAAAGAGA 840
OY 841 TGGAGCTGGGCAATGAGGAAATTTTGTCCGGTTTGGTGTATATCTCTGAGAGAAA 900
   |||||||
DB 841 TGGAGCTGGGCAATGAGGAAATTTTGTCCGGTTTGGTGTATATCTCTGAGAGAAA 900
OY 901 ACCTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGATAGCAATTAATA 954
   |||||||
DB 901 ACCTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGATAGCAATTAATA 954

```

RESULT 3
 AA246759
 ID AA246759 standard; cDNA; 954 BP.

XX AC AA246759;

XX DT 31-MAR-2000 (first entry)

XX DE T. maritima CMCase enzyme encoding cDNA.

XX KW Endoglucanase; carboxymethyl cellulase; cellulose; CMCase; plant biomass;
 KW beta-1,4-glycosidic bond; detergent; textile; guar gum; animal feed;
 KW waste product; fuel; brewing; thermostable; ss.

XX OS Thermotoga maritima.

XX FH key
 XX FT CDS location/Qualifiers
 XX FT 1..954 /tag- a
 XX FT /product- "CMCase"

XX PN US6008032-A.

XX PD 28-DEC-1999.

XX PF 16-OCT-1997; 97US-0951889.

XX PR 23-AUG-1995; 95US-0518615.

XX PA (DIVE-) DIVERSA CORP.

XX PI Mathur EJ, Lam DE;

XX DR WPI: 2000-105335/09.

XX DR P-PSDB: AAY56814.

XX PT Endoglucanase with carboxymethyl cellulase activity from Thermotoga
 XX PT maritima, useful for degrading biomass to fuels and chemicals

XX PS Disclosure; Fig 1A-C; 14pp; English.

XX CC This cDNA encodes a T. maritima endoglucanase with carboxymethyl
 XX CC cellulase activity. The enzyme CMCase is thermostable and used to degrade
 XX CC carboxymethyl cellulose (CMC) or more generally any oligosaccharide that
 XX CC contains beta-1,4-glycosidic bonds, particularly in detergents, textiles,
 XX CC guar gum, animal feed, plant biomass or waste products. The CMCase enzyme
 XX CC is especially used to convert plant biomass to chemicals and fuels, and
 XX CC also for clarification and extraction of juice (e.g. In the brewing
 XX CC industry). CMCase can also be used to raise specific antibodies.

XX SQ Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other;

Query Match 100.0%; Score 954; DB 21; Length 954;
 Best Local Similarity 100.0%; Pred. No. 1.6e-244;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ATGGGTTGATCCTTTTGAAGCAAAATTTGGGAAGGCACTTAATATAGCAAAAT 60
   |||||||
DB 1 ATGGGTTGATCCTTTTGAAGCAAAATTTGGGAAGGCACTTAATATAGCAAAAT 60
OY 61 GCGCTTGAAGCAACCAATGAGGAGACTGGGAGTGTATAAAGATGAGTCTTGAC 120
   |||||||
DB 61 GCGCTTGAAGCAACCAATGAGGAGACTGGGAGTGTATAAAGATGAGTCTTGAC 120
OY 121 ATTATTAAGAAGCCGGTTTCTCATGTTCCGAATTCCAATAGATGAGTACGACGCT 180
   |||||||
DB 121 ATTATTAAGAAGCCGGTTTCTCATGTTCCGAATTCCAATAGATGAGTACGACGCT 180
OY 181 TACGCTTTCCTCTTATTAATCATGATCGCTTCTCAAAAGAGTGGATGAAGTATA 240
   |||||||
DB 181 TACGCTTTCCTCTTATTAATCATGATCGCTTCTCAAAAGAGTGGATGAAGTATA 240
OY 241 AACGAGCCCTGAAAAGAGACTGGCTGTGCTATAAATATTCATCAGTACGAGAGTTA 300
   |||||||

```

```

Db 241 AACGAGCCCTGAAAGAGAGACTGCTGTCTATTAATATTCACCTACAGAGAGTTA 300
QY 301 ATGATGATCCAGAGAAACACAGAGAAATTTCTGCTCTTGGAAACAATTCCTGAT 360
Db 301 ATGATGATCCAGAGAAACACAGAGAAATTTCTGCTCTTGGAAACAATTCCTGAT 360
QY 361 CGTTATTAAGACTATCCCGAAACTCTATTTTGAATTTCTGAATGAACCTCAGCGAAT 420
Db 361 CGTTATTAAGACTATCCCGAAACTCTATTTTGAATTTCTGAATGAACCTCAGCGAAT 420
QY 421 CTCTACCCGAGAAATTTGAATGATGCTCTTGGAGAAAGCTCTAAAGCTTAAGATCAAT 480
Db 421 CTCTACCCGAGAAATTTGAATGATGCTCTTGGAGAAAGCTCTAAAGCTTAAGATCAAT 480
QY 481 GACAAAGAGACACTATATATATATAGGACAGCTGAATGGGGGGGTATATCTCCCTTGA 540
Db 481 GACAAAGAGACACTATATATATATAGGACAGCTGAATGGGGGGGTATATCTCCCTTGA 540
QY 541 AAACGTCTGTCGCAAAATGGGAAAAAATTTCTATATGTTACATTCACCTACTACATCC 600
Db 541 AAACGTCTGTCGCAAAATGGGAAAAAATTTCTATGTTACATTCACCTACTACATCC 600
QY 601 TTGCAATTTACCCATCAGAGAGCTGAGTGGGGAAGATCGAGAAATGGTTGGAGAA 660
Db 601 TTGCAATTTACCCATCAGAGAGCTGAGTGGGGAAGATCGAGAAATGGTTGGAGAA 660
QY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAGAAATTTATAGAGAA 720
Db 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAGAAATTTATAGAGAA 720
QY 721 TGGTCAAAAAAACAACAAAGACCAATTTACATAGTGAATTTGGTCCCTACAGAAAGCT 780
Db 721 TGGTCAAAAAAACAACAAAGACCAATTTACATAGTGAATTTGGTCCCTACAGAAAGCT 780
QY 781 GACCTTGAATCAAGATTAATGAGACCTCTTGTGCTTCGCGAAATGAGAAAGAGAG 840
Db 781 GACCTTGAATCAAGATTAATGAGACCTCTTGTGCTTCGCGAAATGAGAAAGAGAG 840
QY 841 TGGAGCTGGGCACTACTGGAATTTTGTCCGGTTTGGTGTATATGATCTGAGAGAA 900
Db 841 TGGAGCTGGGCACTACTGGAATTTTGTCCGGTTTGGTGTATATGATCTGAGAGAA 900
QY 901 ACCTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGATAGCATTTGAATA 954
Db 901 ACCTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGATAGCATTTGAATA 954

RESULT 4
AADI5243
ID AADI5243 standard: DNA; 954 BP.
XX
AC AADI5243:
XX
DT 15-NOV-2001 (first entry)
XX
DE Thermotoga maritima CMCase enzyme encoding DNA.
XX
KM Carboxymethyl cellulose; CMCase; plant biomass; animal feed; detergent;
KW textile industry; waste treatment; brewing industry; immunogen;
KM fruit juice industry; ds.
XX
OS Thermotoga maritima.
XX
FH Key
FT 1..954 Location/Qualifiers
FT /tag=a
FT /product="T. maritima CMCase enzyme"
PN US6245547-B1.
PD 12-JUN-2001.
XX

```

```

PF 27-DEC-1999; 99US-0472857.
XX
XX 23-AUG-1995; 95US-0518615.
PR 16-OCT-1997; 97US-0951889.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Mathur EJ, Iam DE:
XX WPI: 2001-528380/58.
DR P-PSDB: AA08543.
XX
XX Polynucleotide isolated from thermophile eubacteria useful for
PT degradation of cellulose encodes an enzyme having a fully defined amino
PT acid sequence.
XX
XX Claim 3; Fig 1; 14pp; English.
XX
XX The present sequence is a DNA coding for a thermostable enzyme, an
CC endoglucanase having carboxymethyl cellulose activity referred to as
CC (CMCase). The CMCase enzyme is useful for the degradation of cellulose,
CC for conversion of plant biomass into fuels and chemicals, in the
CC textile industry, in detergents, in animal feed, in waste treatment,
CC in fruit juice/brewing industry for clarification and extraction of
CC juices, to generate probes for identifying sequences encoding similar
CC enzymes from other organisms, as immunogen to produce monoclonal or
CC polyclonal antibodies.
XX
XX Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other:
SQ
Query Match 100.0%; Score 954; DB 22; Length 954;
Best Local Similarity 100.0%; Pred. No. 1,6e-244;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTGTTGATCCCTTTTGAAGGAAACAATATTTGGAGAGCATTAATATAGAAAT 60
Db 1 ATGGGCTTTGATCCCTTTTGAAGGAAACAATATTTGGAGAGCATTAATATAGAAAT 60
QY 61 GCGCTTGAAGCACCAATATGAGGAGAGCTGGGAGTGTGATTAAGATGAGTCTTGAC 120
Db 61 GCGCTTGAAGCACCAATATGAGGAGAGCTGGGAGTGTGATTAAGATGAGTCTTGAC 120
QY 121 ATTATTAAGAAAGCCGGTTTCTCTCATGTTCGAATTTCCAAATTAAGATGAGTACGAC 180
Db 121 ATTATTAAGAAAGCCGGTTTCTCTCATGTTCGAATTTCCAAATTAAGATGAGTACGAC 180
QY 181 TACGCTTTCCTCCTTATTAATATCATGATCGCTTCTTCAAAAGAGTGAATGATGATA 240
Db 181 TACGCTTTCCTCCTTATTAATATCATGATCGCTTCTTCAAAAGAGTGAATGATGATA 240
QY 241 AACGAGCCCTGAAAGAGAGAGCTGCTGTCTATTAATATTCATCTACGAGAGTTA 300
Db 241 AACGAGCCCTGAAAGAGAGAGCTGCTGTCTATTAATATTCATCTACGAGAGTTA 300
QY 301 ATGAATGATCCAGAGAAACACAGAGAAATTTCTGCTCTTGGAAACAATTCCTGAT 360
Db 301 ATGAATGATCCAGAGAAACACAGAGAAATTTCTGCTCTTGGAAACAATTCCTGAT 360
QY 361 CGTTATTAAGACTATCCCGAAACTCTATTTTGAATTTCTGAATGAACCTCAGCGAAT 420
Db 361 CGTTATTAAGACTATCCCGAAACTCTATTTTGAATTTCTGAATGAACCTCAGCGAAT 420
QY 421 CTCTACCCGAGAAATTTGAATGATGCTCTTGGAGAAAGCTCTAAAGCTTAAGATCAAT 480
Db 421 CTCTACCCGAGAAATTTGAATGATGCTCTTGGAGAAAGCTCTAAAGCTTAAGATCAAT 480
QY 481 GACAAAGAGACACTATATATATATAGGACAGCTGAATGGGGGGGTATATCTCCCTTGA 540
Db 481 GACAAAGAGACACTATATATATATAGGACAGCTGAATGGGGGGGTATATCTCCCTTGA 540
QY 541 AAACGTCTGTCGCAAAATGGGAAAAAATTTCTATATGTTACATTCACCTACTACATCC 600
Db 541 AAACGTCTGTCGCAAAATGGGAAAAAATTTCTATGTTACATTCACCTACTACATCC 600

```


Db 986 ATTTGGAAATTTTGTGACAGATTGGCATATAGATGATGCTCAAACTGGATCGAAC 1045
 QY 914 ATCTTTAGAGCTTTAATAGAGGAG 940
 Db 1046 CATTGGCAACGCTGTGGTTGGCAGAG 1072

RESULT 6
 AAT93690
 AAT93690 standard: DNA: 1083 BP.

AC AAT93690:
 XX 12-MAR-1998 (first entry)
 XX Thermotoga ocl/4V endoglucanase encoding DNA.
 DE Thermotoga ocl/4V endoglucanase encoding DNA.
 XX Glycosidase; thermostable; textile; food processing; pharmaceutical;
 KM detergent; baking; industry; Thermococcus; staphylothermus;
 KW Pyrococcus; glucose; soluble oligosaccharide; endoglucanase; ss.
 OS Thermotoga sp.

XX Key Location/Qualifiers
 FT 1.1083
 FT /tag- a
 FT /product- Endoglucanase

XX MO9725417-A1.
 XX 17-JUL-1997.
 XX 10-JAN-1997: 97WO-US00092.
 XX 13-SEP-1996: 96US-0712612.
 XX 11-JAN-1996: 96US-0583787.
 XX (RECO-) RECOMBINANT BIOCATALYSIS INC.

XX Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
 DR WPI: 1997-37258/34.
 DR P-PSDB: AAW34566.

PT New thermostable glycosidase(s) - from Thermococcus, Staphylothermus
 PT and Pyrococcus, used in the textile, food processing,
 PT pharmaceutical, detergent and baking industries

PS Claim 4; Fig 13; 82pp; English.

CC The present sequence encodes endoglucanase isolated from Thermotoga.
 CC The enzyme or its encoding nucleic acid sequence is used for generating
 CC glucose from soluble oligosaccharides. The enzyme can be used in the
 CC food processing, pharmaceutical, textile, detergent and baking
 CC industries. The enzyme is also used to treat lactose intolerance, as a
 CC diagnostic reporter molecule, in corn wet milling or in the fruit juice
 CC industry. The enzymes can be used to hydrolyse guar gum to remove
 CC non-reducing terminal mannose residues. The nucleic acids encoding the
 CC enzyme may be used to generate probes to identify similar sequences.

XX Sequence 1083 BP; 368 A; 191 C; 239 G; 285 T; 0 other;

Query Match 24.4%; Score 232.6; DB 18; Length 1083;
 Best Local Similarity 53.9%; Pred. No. 4.7e-52;
 Matches 500; Conservative 0; Mismatches 424; Indels 3; Gaps 1;

QY 14 CTTTGAAGGAAACAATATTTGGAGAGGACATTATATAGCAATCGCTTGAACAC 73
 Db 149 CATTGATATCAACAAATATGTAAGAGAGTAATATTTGAAATGCTTTAGAACCTC 208
 QY 74 CAATATGAGGAGACTGGGGAGTGTGATAAAGTGTCTTTCGACATTATAAAGAG 133

Db 209 CTTTGAAGGAGACTGGGAGTAGAATGAGCAATGATATTTGAGATATAACAAA 268
 QY 134 CCGGTTTCTTCATGTTGGATTTCCAAATAGATGAGTACGACGCTTACGCTTTCCTC 193
 Db 269 GGGGATTTGATTCCTGTTAGATCCCAATGATGTCAGACATATATCCGAAAGCCAC 328
 QY 194 CTTTAAATATCATGATGCTGCTTTCAAAAGAGTGCATGATTAACGAGCCCTGA 253
 Db 329 CATATGATATTCAGAGCAATTTCTCGAAGAGATTAACCATGTTGTGATAGGCTCTTG 388
 QY 254 AAAGAGGAGCTGGCTGTTGCTATTAATATTCATCACTACGAGGATTAATGATTCAG 313
 Db 389 ACAATATTTAATACAGTATCAATACGACACATTTTGAAGAACTTATCAAGAACCGG 448
 QY 314 AAGAACACAGAAAGATTTCTGCTTTGGAAACAAATGCTGATGCTTATTAAGACT 373
 Db 449 ATMAATACGGCGATGTTTGGGAAATTTGGAGACAGATTGCAAAATTTCTTTAAAGATT 508
 QY 374 ATCCGAAACTCTATTTTGGAAATTTGAAATTCGATGAACTCAAGAAATCTTACTCGGAAA 433
 Db 509 ACCCGAAATATGTTCTTTGAAATCTCAACGAGCGCTCAGAACTTGACAGCTGAAA 568
 QY 434 AATGGAATGAGTCTGTTGAGAGAGCTCAAAAGTTATTAAGTCAATGACAAAAGCACA 493
 Db 569 AATGGAAGCGCTTATTCAAAAGTGTCAAAAGTTATCAGGAGAGCAATCCAAACCGGA 628
 QY 494 CTATATATTTAGGACAGCTGATGAGGGGGGTATATCTGCCCTTGAAGAACTGTCTGCC 553
 Db 629 TTGTCATTATGATGCTCCCAAACTGCGACACATATAGCGAGTGAAGTCTTAATATAG 688
 QY 554 CAATATGGAAGAAATTTCTATGATTAATGATTAATGATTAATGATTAATGATTAATG 613
 Db 689 TCAAA--CGACAAAGCATCATTTGTTCTTCCTTCATTAAGAACTTCAAAATTCACAC 745
 QY 614 ATCAAGAGCTGAGTGGGGGAGAGATCTGAGAAATGTTGGGAGAAAGTGGGATCTC 673
 Db 746 ATCAGAGGTCCTCAATGGGTTAATCCCATCCACCTGTTAGGGTTAAGTGAAGTGGCAGG 805
 QY 674 CAGATGATCAGAAAGATTTGATGAAGATTAATGATTAATGATTAATGATTAATGATTA 733
 Db 806 AATGGAATTTAACAATCAAGATGATTCATTAATGATGATGATGATGATGATGATGAT 865
 QY 734 ACAAAAGACCAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 793
 Db 866 ATACGATACCAATTTCTTGGTGAATGATGATGATGATGATGATGATGATGATGATGAT 925
 QY 794 GATTAATGAGACCTTCCTTTCGTTTCGGAATGAGAAAGGATGAGGATGAGGAT 853
 Db 926 GGGTTAAGTGCACCGAAAGTGTGAGAAAGTGTGAGAAATTTGATTTTCATACGGCT 985
 QY 854 ACTGGAATTTTGTCCGTTTGTGTTGATGATGATGATGATGATGATGATGATGATGAT 913
 Db 986 ATGCGAATTTTATGAGATTTGGCATATGATGATGATGATGATGATGATGATGATGATG 1045
 QY 914 ATCTTTAGAGCTTTAATAGAGGAG 940
 Db 1046 CATTGGCAACGCTGTGGTTGGCAGAG 1072

RESULT 7
 AAV29477
 AAV29477 standard: cDNA to mRNA; 1826 BP.

XX AAV29477;
 AC 21-SEP-1998 (first entry)
 DT Orpinomyces cellulase CelB cDNA.
 DE Orpinomyces cellulase CelB cDNA.
 KW Cellulase; endoglucanase; CelB gene; ds.
 XX Orpinomyces sp. strain FC-2.

Key Location/Qualifiers
 CDS 69..1484
 /tag= a
 /note= "Claim 5"

MO9814597-A1.
 09-APR-1998.
 03-OCT-1997; 97WO-US18008.
 04-OCT-1996; 96US-0027883.
 (UYGE-) UNIV GEORGIA RES FOUND INC.
 PI Chen H, Li X, Ljungdahl LG;
 PI MPI; 1998-240096/21.
 DR P-PSDB; AAM56742.

New recombinant DNA encoding Orpinomyces cellulase protein - useful for, e.g. producing recombinant Orpinomyces cellulase in host cell

Claim 5: Page 38-40; 69pp; English.

This cDNA clone includes a claimed coding region for cellulase celB (see AAM56742) of the anaerobic bovine rumen fungus Orpinomyces sp. PC-2. It was obtained by screening a PC-2 cDNA library for clones active on remazol brilliant blue-carboxymethylcellulose. The encoded cellulase has endoglucanase, but not cellobiohydrolase, activity. CelA and celC genes (see AAV29472-73), also obtained from Orpinomyces sp. PC-2, encode cellulases having both activities (see AAM56738-39). Recombinant DNA molecules encoding Orpinomyces cellulase proteins are claimed, as well as recombinant cells selected from Saccharomyces cerevisiae, Escherichia coli, Aspergillus, Trichoderma reesei, Pichia, Penicillium, Streptomyces or Bacillus, and a method for producing recombinant cellulase by culturing these host cells.

Sequence 1826 BP; 654 A; 236 C; 313 G; 623 T; 0 other:

Query Match 7.7%; Score 73; DB 19; Length 1826;
 Best Local Similarity 50.4%; Pred. No. 2.1e-09;
 Matches 212; Conservative 0; Mismatches 200; Indels 9; Gaps 1;

107 ATGAGTTCTTCGACATTATAAAGAGCCGGTTCTCTCATGTCGATTCGATTAAGAT 166
 277 AAGAGCTTTACTATATAACTAGTATGATCTGGTTTCAACACTTTCCGATTCGATTAAGAT 336
 167 GGAGTACGACGCGTTACGCGTTCTCTTATAAATCATGATCGCTCTTCAAAAGAG 226
 337 GGAGTGGTCTATTTGGTATGCTGCTGACTATTAATTTAGTATGTTGGATGAAGAG 396
 227 TGGATGATGATTAAGAGAGCCCTTAAGAGAGGACTGCTGTTGATTAATTAATTCATC 286
 397 TTTATGATGATTTGCTGTTTATGCTTAAACAGTGGTGTATGATTAATTAATTCATC 456
 287 ACTAGC-----AGGATTAATGATGATGATGATGATGATGATGATGATGATGATG 337
 457 ATGAAACTGGAATTTATGCTTTCCAAAGAAATTTAGAGAGTGCCTTAAGTTG 516
 338 CTTCTTGAAGAAATTTGCTGATGCTTATAAGACTATCCGAAATCTATTTTGAAG 397
 517 CCATCTGGAAGAAATTTGCTGATTTGATGATTAATGATGATTAATTTTGAAG 576
 398 TTTCTGATTAACCTCAAGAAATCTTATCCGAAAGAAATGAAATGATGATGATGATG 457
 577 GAATGATTAACCAAGAAAGTTGGGATCCAGCTGAATGATGATGATGATGATGATG 636
 458 CTTCTAAAGTTTAAGATGATTAAGATGATGATGATGATGATGATGATGATGATG 517
 637 GTTGAATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 696

Qy 518 G 518
 Db 697 G 697

RESULT 8
 AAV37413
 ID AAV37413 standard; cDNA to mRNA; 1826 BP.
 XX
 AC AAV37413;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Orpinomyces cellulase CelB cDNA.
 XX
 KW Cellulase; endoglucanase; CelB gene; ds.
 XX
 OS Orpinomyces sp. strain PC-2.
 XX
 OS MPI; 1998-240096/21.
 XX
 FH Key Location/Qualifiers
 FT CDS 69..1484
 FT /tag= a
 FT /note= "Claim 5"

MO9814597-A1.
 09-APR-1998.
 03-OCT-1997; 97WO-US18008.
 04-OCT-1996; 96US-0027883.
 (UYGE-) UNIV GEORGIA RES FOUND INC.
 PI Chen H, Li X, Ljungdahl LG;
 PI MPI; 1998-240096/21.
 DR P-PSDB; AAM56742.

New recombinant DNA encoding Orpinomyces cellulase protein - useful for, e.g. producing recombinant Orpinomyces cellulase in host cell

Claim 5: Page 38-40; 69pp; English.

This cDNA clone includes a claimed coding region for cellulase celB (see AAM56742) of the anaerobic bovine rumen fungus Orpinomyces sp. PC-2. It was obtained by screening a PC-2 cDNA library for clones active on remazol brilliant blue-carboxymethylcellulose. The encoded cellulase has endoglucanase, but not cellobiohydrolase, activity. CelA and celC genes (see AAV29472-73), also obtained from Orpinomyces sp. PC-2, encode cellulases having both activities (see AAM56738-39). Recombinant DNA molecules encoding Orpinomyces cellulase proteins are claimed, as well as recombinant cells selected from Saccharomyces cerevisiae, Escherichia coli, Aspergillus, Trichoderma reesei, Pichia, Penicillium, Streptomyces or Bacillus, and a method for producing recombinant cellulase by culturing these host cells.

Sequence 1826 BP; 654 A; 236 C; 313 G; 623 T; 0 other:

Query Match 7.7%; Score 73; DB 19; Length 1826;
 Best Local Similarity 50.4%; Pred. No. 2.1e-09;
 Matches 212; Conservative 0; Mismatches 200; Indels 9; Gaps 1;

107 ATGAGTTCTTCGACATTATAAAGAGCCGGTTCTCTCATGTTGATTCGATTAAGAT 166
 277 AAGAGCTTTACTATATAACTAGTATGATCTGGTTTCAACACTTTCCGATTCGATTAAGAT 336
 167 GGAGTACGACGCGTTACGCGTTCTCTTATAAATCATGATCGCTCTTCAAAAGAG 226
 337 GGAGTGGTCTATTTGGTATGCTGCTGACTATTAATTTAGTATGTTGGATGAAGAG 396
 227 TGGATGATGATTAAGAGAGCCCTTAAGAGAGGACTGCTGTTGATTAATTAATTCATC 286

Seq	Sequence	1532 BP;	494 A;	239 C;	294 G;	505 T;	0 other;
QY	Query Match	7.2%;	Score 68.6;	DB 21;	Length 1532;		
Db	Best Local Similarity	53.2%;	Fred. No. 3e-08;				
Matches 1/4;	Conservative	0;	Mismatches 144;	Indels 9;	Gaps 1;		
OY	125 TAAAGAGACCGGTTCTCTCATGTTCCGAAATTCACAAATAGATGAGTACGACGCTTTACG	184					
Db	268 TAAAGATCAAGSTTTTAAATGTTTCCGTAATCCAAACCACTTGACGCGTCACTTGGTA	327					
OY	185 CGTTCTCCTTTTAAATCATGATGATCGCTTCTTCCAAAGATGTGATGAATGATAACG	244					
Db	328 ATGCTCCAGATTCACAAATATGACTGATGTTGGATGAGAAAGCTTCATGACTTGTGATT	387					
OY	245 GAGCCCTGAAAAGAGCAGCTGCGTGTGCTAATAATATTCATCAGTACAGAGC-----	297					
Db	388 ATGCTCTTAACACTGGACAGCTAGCTATCTTAAATATTCACCATGAAATGGAATTTATG	447					
OY	298 --TTAATGATGATGATCCAGAAACACACAGAAAGAACTTCTGCTTTGCAACAAATTTG	355					
Db	448 CTTCCTCTTAACAATTTTACAAAGCAAAACCAATTTTACTGCTGATCTGGAACAAATTTG	507					
OY	356 CTGATCGTATTAAGACTATCCGAAACACTATTTTGTGAATTCGTAATGTAACCTCAGC	415					
Db	508 CTGCTGAATTCGCTAAGTATGATGATGATTTTGAATTTGGAAGGTATGAAATGAAACAGAA	567					
OY	416 GAAATCTTACTCCGAGAAATGGAATG 442						
Db	568 AGCTTGATCATCCAAATGAATGGAATG 594						
RESULT 10							
AAQ55036							
ID	AAQ55036 standard; cDNA; 4098 BP.						
XX	AAQ55036;						
AC							
XX	21-JUL-1994 (first entry)						
DT							
XX	Sequence of celd cDNA in clone pcNP4.						
DE							
XX	Cellulase: celd; pcNP4; anaerobic rumen; ss.						
KM							
XX	Neocallimastix patriciarum.						
OS							
XX							
FH	Key	Location/Qualifiers					
FT	CDS	1..2400					
FT		/*tag= a					
FT		/product= domain I and II					
FT	CDS	1..111					
FT		/*tag= b					
FT		/product= N-terminus of beta-galactosidase alpha-peptide					
FT		/note= "derived from pbluescript SK-vector"					
FT	misc_difference	112..124					
FT		/note= c					
FT		/note= "derived from oligo linker"					
FT	CDS	2468..4098-					
FT		/*tag= d					
FT		/product= domain III					
PN	WO9400578-A.						
XX							
PD	06-JAN-1994.						
XX							
PF	24-JUN-1993;	93WO-AU00307.					
XX							
PR	24-JUN-1992;	92AU-0003096.					
XX							
PA	(CSIR) COMMONWEALTH SCI & IND RES ORG.						
XX							
PI	Aylward JH, Gobius KS, Orpin CG, Xue GP;						

DR WPI: 1994-026214/03.
DR P-PSDB; AAR47496, AAR49102.
XX
XX Cloning of cellulase clones from anaerobic rumen - by isolating
PT mRNA from culture, converting to cDNA etc. fungi, producing
PT enzymes useful in food processing etc., and DNA for modifying
PT rumen or silage bacteria
XX
PS Claim 9; Page 39-40; 71pp; English.
XX
CC Clone pcn4.1 encodes celd. It was derived as follows. N.
CC patriciarum was used to construct a cDNA library in ZAP1 (in E.
CC coli). Transformants were selected for enzyme activity using
CC selective media. 11 colonies were positive, and of these 10 had the
CC same restriction pattern, and the longest of these was designated
CC celd (pcn4.1) (AA055036) (AAR47496, R49102). A similar process was used
CC to isolate the xylanase clone pnx-tac (AA055037, R47497). An enzyme
CC composition contg. celd and xylanase is claimed.
XX
XX Sequence 4098 BP; 1386 A; 598 C; 835 G; 1279 T; 0 other;
SQ
Query Match 6.6%; Score 63.2; DB 15; Length 4098;
Best Local Similarity 50.5%; Pred. No. 1.2e-06;
Matches 186; Conservative 0; Mismatches 173; Indels 9; Gaps 1;
OY 99 GATMAAGAGAGTCTTCGACATTTAAAGAGCGGTTCTCTCATGTTGGAATTC 158
DB 2632 GACTACTGAAGATATGTTCAGGTTTAAATGATACCAATTTAATGTTTCGTAATTC 2691
OY 159 AATAAGATGAGTACGACGCTTACCGGTTCTCTTATAAATCATGATCGCTTCT 218
DB 2692 AACTACTTGGTCTGTCACCTTCGTAAGCTCCAGATTTACAGATTAATGAAATGTT 2751
OY 219 CAAAGAGTGGATGATGATTAACGAGCCCTGAAAGAGGACCTGGCTTCTATAA 278
DB 2752 AAGAGAGTTCATGAAATTTGTTATTCATCAAGATGAGCTTTCGTTATCTTAA 2811
OY 279 TATTCATCACT-----ACGAGAGTATGATGATCCAGAACACAGAGAAAG 329
DB 2812 TCTTCACATGAAGATGAGAACCATGCTTCTCTGAACCTTGACACATGCCAGAGAAAT 2871
OY 330 ATTTCTTCTCTTGGAAACAATTCGTGATGCTTATAAGACTATCCGAAACTCTATT 389
DB 2872 TTTAGAAGATTTGGTCTCAATTCGTGAAGATTTAAGATTAATGATGATGACACTTAAT 2931
OY 390 TTTTGAATTTCTGAATGAACCTCAGCAAAATCTTACTCCGGAATAATGGAATGACTGCT 449
DB 2932 TTTTGAAGATTTAAGCAACCAAGAAAGATGATCTCCAGTTGAATGGCTGGTGA 2991
OY 450 TGAGGAAG 457
DB 2992 TCAGGAAG 2999
RESULT 11
AAF83293.
ID AAF83293 standard; cDNA: 1170 BP.
XX
XX AAF83293;
AC
XX 09-JUL-2001 (first entry)
DT
XX
DE P. rhizinflata egla protein catalytic domain encoding cDNA.
XX
XX Cellulase; egla: fungus; polysaccharide; glycosidic linkage; enzyme;
KW animal feed; alcohol; fermentation; catalytic domain; ss.
XX
OS Pirmoyces rhizinflata.
XX
XX
FH Key Location/Qualifiers
FT 1.1170
FT CDS
FT /tag- a
FT /product- "egla catalytic domain"

XX
XX US6222028-BI.
XX
XX
XX 24-APR-2001.
XX
XX
XX 15-OCT-1999; 99US-0419459.
XX
XX 15-OCT-1999; 99US-0419459.
XX
XX (SINT-) ACAD STINICA.
XX
XX
XX Liu J, Cheng K, Tsai C, Chang C;
XX
XX WPI: 2001-307598/32.
XX
XX P-PSDB; AAB62468.
XX
XX
XX New isolated nucleic acid encoding a cellulase enzyme of fungus
PT Pirmoyces rhizinflata, for producing cellulase that hydrolyzes
PT polysaccharides containing beta-1,3 or beta-1,4 glycosidic linkages
XX
XX
PS Claim 3; Columns 17-20; 16pp; English.
XX
XX
CC The invention relates to a cellulase enzyme, egla, isolated from the
CC fungus Pirmoyces rhizinflata. The egla polypeptide hydrolyzes a
CC polysaccharide containing a beta-1,3' or beta-1,4' glycosidic linkage.
CC The egla polynucleotide is useful for producing cellulase enzymes that
CC can hydrolyze the glycosidic linkages in polysaccharides such as
CC cellulose. The enzyme is also useful in a number of industrial
CC applications where breaking down biomass is beneficial. Cellulase is also
CC useful as a supplement in animal feed to decrease the production of fecal
CC waste by increasing the digestibility of the feed. Cellulase is also
CC useful to increase the efficiency of alcoholic fermentations by
CC converting undigestible biomass into fermentable sugars. The present
CC sequence represents a cDNA encoding the catalytic domain of egla
CC polypeptide.
XX
XX
SQ Sequence 1170 BP; 385 A; 182 C; 247 G; 356 T; 0 other;
Query Match 6.6%; Score 62.8; DB 22; Length 1170;
Best Local Similarity 49.7%; Pred. No. 9.5e-07;
Matches 194; Conservative 0; Mismatches 187; Indels 9; Gaps 1;
OY 99 GATMAAGAGAGTCTTCGACATTTAAAGAGCGGTTCTCTCATGTTGGAATTC 158
DB 228 GACTACTGAAGATATGTTCAGGTTTAAATGATACCAACCAATTTAATGCTTCGTAATTC 287
OY 159 AATAAGATGAGTACGACGCTTACCGGTTCTCTTATAAATCATGATCGCTTCT 218
DB 288 AACTACTTGGTCTGTCACCTTCGTAAGCTCCAGATTTAAGATTAATGAAATGTT 347
OY 219 CAAAGAGTGGATGATGATTAACGAGCCCTGAAAGAGGACCTGGCTTCTATAA 278
DB 348 AAGAGAGTTCATGAAAGTTGTTTATTCATCAAGAGACGAGCAATTTGTTTATAA 407
OY 279 TATTCATCACTAG-----AGGAGTTAATGATGATCCAGAACACAGAGAAAG 329
DB 408 TCTTCATCATGAAGATGAGATCATGCTTCTCTGAACCTTGTAATGACCAAGAAAT 467
OY 330 ATTTCTTCTCTTGGAAACAATTCGTGATGCTTATAAGACTATCCGAAACTCTATT 389
DB 468 TTTAGAAAAGATCTGCTCAATTCGTGAAGATTTAAGATTAATGATGATGACACTTAAT 527
OY 390 TTTTGAATTTCTGAATGAACCTCAGCAAAATCTTACTCCGGAATAATGGAATGACTGCT 449
DB 528 CTTCGAAAGATTTAAGCAACCAAGAAAGATGATCTCCAGTTGAATGGCTGGTGA 587
OY 450 TGAGGAAGCTTAAGATTATAAGATCAAT 479
DB 588 TCAGGAAGTTGGATGCTGTTAATGCTAT 617
RESULT 12
AAF83296

[illegible]

QY	279	TATTCATCAGCTACTG-----ACAGTCTTAAATGAATGCACGAAACACACAGAAGAAG	329
Dd	420	TCCTTCATTCAAGAACCTGGAAATCATGCCCTTCTCTGAAGACTTTGATRCACCCAGAAGAAAT	479
QY	330	ATTTCCTTGCTCTTTGGAAAACAATAATGTGTCGATCGTTAATAAGACTATCCGAAACTCTATT	389
Dd	480	TTTTGAAAAAATATCTGCTCTCAAAATTTGCTGGAAGAAATTAAAGATTATGATCAACTTAAT	539
QY	390	TTTTGAAATTCGTGAATGAACTCACGGAATCTTACTCCGSAAAAAAGAAATGAACATCGCT	449
Dd	540	CTTTCGAAGAGATTAAACGAAACCAGAAAGAAATGATATCTCCAGTTGAATGAGACTGTGGTGA	599
QY	450	TGAGSAGAGCTCTAAAGTTATAAGATCAAT	479
Dd	600	TCAGAAAGGTTGGGATGCTGTTATATGCTAT	629
 RESULT 13 AAF83292			
ID	AAF83292 standard; cdNA; 2110 BP.		
XX	AAF83292;		
AC			
DT	09-JUL-2001	(first entry)	
XX			
DE	P. rhizinflata egla protein encoding partial cdNA.		
XX			
KM	Cellulase; egla; fungus; polysaccharide; glycosidic linkage; enzyme;		
KW	animal feed; alcohol; fermentation; ss.		
OS	Ptomyces rhizinflata.		
XX			
XX			
FH	Key	Location/Qualifiers	
FT	CDS	3..1751	
FT		/*tag= a	
FT		/product= "partial egla"	
XX			
PN	US6222028-B1.		
XX			
PD	24-APR-2001.		
XX			
PF	15-OCT-1999;	99US-0419459.	
PR	15-OCT-1999;	99US-0419459.	
XX			
PA	(SINI-) ACAD SINICA.		
PI	Liu J, Cheng K, Tsai C, Chang C;		
XX			
DR	WPI: 2001-307598/32.		
XX	P-PsDB: AAB62467.		
PT	New isolated nucleic acid encoding a cellulase enzyme of fungus		
PT	ptomyces rhizinflata, for producing cellulase that hydrolyses		
XX	polysaccharides containing beta-1,3 or beta-1,4 glycosidic linkages		
PS	Disclosure; Columns 3-6; 16pp; English.		
XX			
CC	The invention relates to a cellulase enzyme, egla, isolated from the		
CC	fungus ptomyces rhizinflata. The egla polypeptide hydrolyzes a		
CC	polysaccharide containing a beta-1,3' or beta-1,4' glycosidic linkage.		
CC	The egla polynucleotide is useful for producing cellulase enzymes that		
CC	can hydrolyze the glycosidic linkages in polysaccharides such as		
CC	cellulose. The enzyme is also useful in a number of industrial		
CC	applications where breaking down biomass is beneficial. Cellulase is also		
CC	useful as a supplement in animal feed to decrease the production of fecal		
CC	waste by increasing the digestibility of the feed. Cellulase is also		
CC	used to increase the efficiency of alcoholic fermentations by		
CC	converting undigestible biomass into fermentable sugars. The present		
CC	sequence represents a cDNA insert in plasmid pPr2301-10, the open reading		
CC	frame of which encodes for a partial egla polypeptide.		
XX			
Sequence	2110 BP; 770 A; 265 C; 413 G; 662 T; 0 other:		

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 09:20:22 ; Search time 1110 Seconds
(without alignments)
13919.364 Million cell updates/sec

Title: US-09-880-729a-1

Perfect score: 954
Sequence: 1 atgggtgtgacatccttga.....gagagatagcatgataaa 954

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estchum:*
3: em_estin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	50.2	5.3	1101	17	CNS0039L
c 2	48.8	5.1	1101	17	CNS0182P
c 3	48.6	5.1	1101	17	CNS0039G
c 4	46.8	4.9	1101	17	CNS0006J
c 5	46.8	4.9	1101	17	CNS0106X
c 6	45.8	4.8	1204	17	CNS016E2

c 7	45.4	4.8	1101	17	CNS00LVS	AL078812 Drosophila
c 8	44.8	4.7	1031	17	CNS00CF2	AL059199 Drosophila
c 9	44	4.6	915	17	CNS015YX	AL106083 Drosophila
c 10	43.8	4.6	996	17	CNS00FHH	AL071063 Drosophila
c 11	43.2	4.5	1001	17	CNS01400	AL103554 Drosophila
c 12	43	4.5	621	17	AQ236641	AQ236641 RPII11-69
c 13	43	4.5	943	17	CNS002IM	AL097720 Drosophila
c 14	43	4.5	1101	17	CNS017KE	AL108152 Drosophila
c 15	42.6	4.5	942	17	CNS04HHV	AL1290956 Tetradon
c 16	42.4	4.4	740	14	BU011716	BU011716 OCJ16P21
c 17	42.4	4.4	775	17	CNS004NP	AL054294 Drosophila
c 18	42.2	4.4	438	17	AQ882597	AQ882597 HS_5407_B
c 19	42.2	4.4	1101	17	CNS00YXR	AL097005 Drosophila
c 20	42	4.4	851	17	CNS015CH	AL105275 Drosophila
c 21	42	4.4	1028	14	BO921957	BO921957 AGENCOURT
c 22	41.8	4.4	424	17	CNS011AO	AL100026 Drosophila
c 23	41.8	4.4	997	17	CNS005TE	AL060767 Drosophila
c 24	41.4	4.3	1101	17	CNS00L8K	AL068205 Drosophila
c 25	41.4	4.3	1141	17	CNS07206	AL425596 Clone BA0
c 26	41.2	4.3	632	17	A2337849	A2337849 IM0068L10
c 27	41.2	4.3	864	17	CNS01RDZ	AL157287 Anopheles
c 28	41.2	4.3	1101	17	CNS0181N	AL108773 Drosophila
c 29	41	4.3	908	17	CNS01678	AL106382 Drosophila
c 30	41	4.3	1101	17	CNS0167I	AL106896 Drosophila
c 31	40.8	4.3	401	14	BO576938	BO576938 PESTOab1
c 32	40.8	4.3	402	14	BO739669	BO739669 PESTOab4
c 33	40.8	4.3	471	13	BI815400	BI815400 PESTOab1
c 34	40.8	4.3	552	14	BO596856	BO596856 PESTOab2
c 35	40.8	4.3	595	14	BO633359	BO633359 PESTOab4
c 36	40.8	4.3	601	14	BO452215	BO452215 PESTOab9
c 37	40.6	4.3	672	17	BH605324	BH605324 BOGKX91TR
c 38	40.6	4.3	914	9	AU234675	AU234675 AU234675
c 39	40.4	4.2	443	9	AU006638	AU006638 AU006638
c 40	40.4	4.2	951	17	CNS015OI	AL105780 Drosophila
c 41	40.2	4.2	928	17	CNS0102F	AL098433 Drosophila
c 42	40.2	4.2	1080	17	CNS00DEP	AL069494 Drosophila
c 43	40.2	4.2	1192	17	CNS03Y36	AL265803 Tetradon
c 44	40	4.2	439	12	BF530485	BF530485 602071726
c 45	40	4.2	510	9	AJ489140	AJ489140 AJ489140

ALIGNMENTS

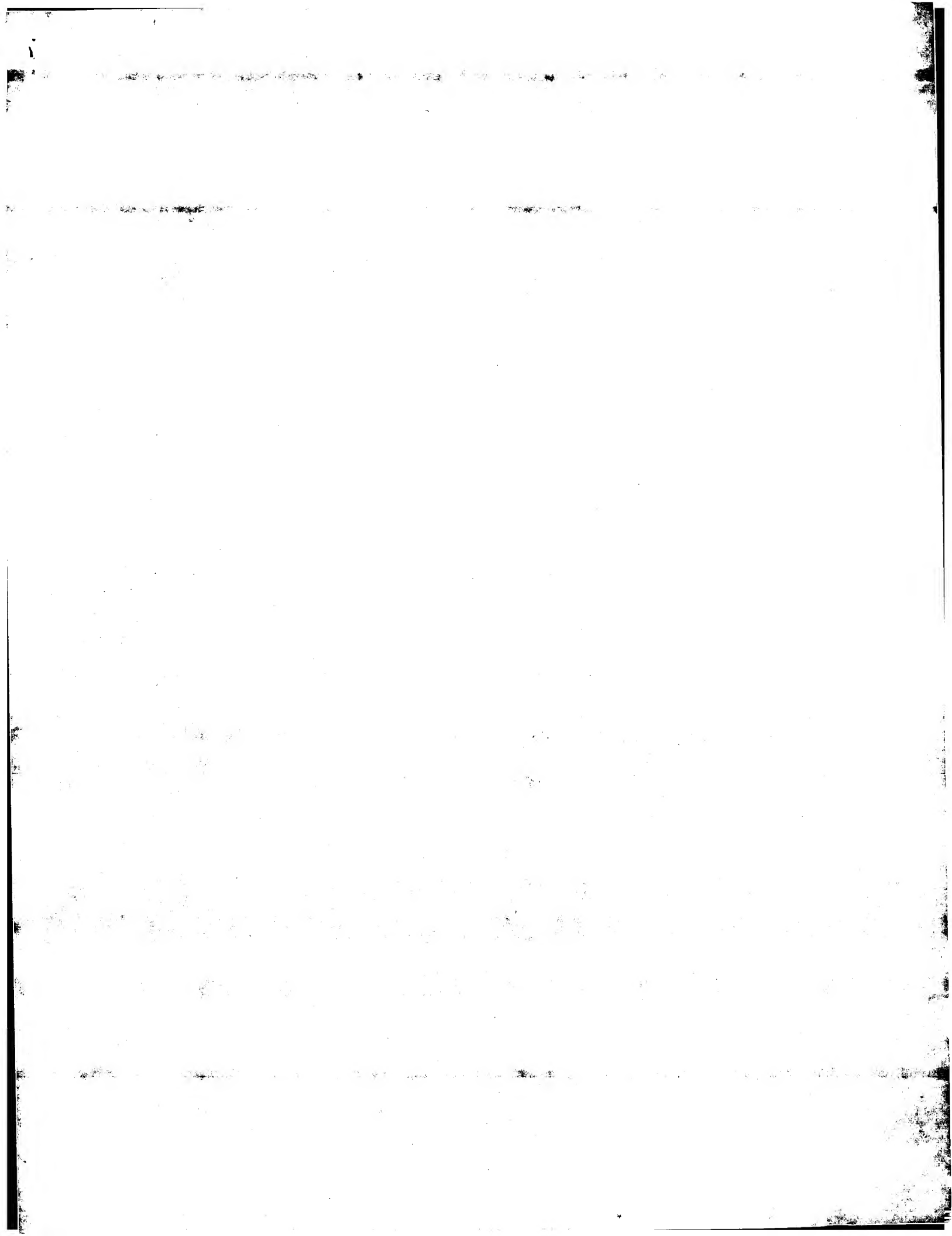
RESULT 1
LOCUS CNS0039L/c 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION .Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR08110 of RPII-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063926 GI:4941783
VERSION AL063926
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPII-98 and was constructed by partial

RESULT 5	CNS0106X	1101 bp	DNA	linear	GSS 26-JUL-1999
LOCUS	CNS0106X				
DEFINITION	Drosophila melanogaster genome survey sequence 17 end of BAC BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL098595				
VERSION	AL098595.1				
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)				
COMMENT	- Web : www.genoscope.cns.fr - Determination of this BAC-end and sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.				
FEATURES	Location/Qualifiers				
source	1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_1b="DrosBAC" /plasmid="pBeloBAC11" /note="end : 17"				
BASE COUNT	258 a 107 c 60 g 175 t 501 others				
ORIGIN					
Query Match	4.9%; Score 46.8; DB 17; Length 1101;				
Best Local Similarity	17.0%; Pred. No. 1;				
Matches	74; Conservative 191; Mismatches 170; Indels 1; Gaps 1				
QY	516 ATGGGGGGTATATCTGCCCCGTTGAAAAACGCTGCTGTCGCCAAATGGGAAAAAATTCCTAT	575			
DB	654 AKATKKKARDDKDTAKDAMKMWKDRATKKAKKAKAKAKAKAAATATGDKARAKAKAKAK	713			
QY	576 AGTTACATTCACCTACTACATCATCTTTCGATTATTCACCATCAAGAGAGCTAGTGGGTGA	635			
DB	714 ATRRKARARMDTATWTATADAAADKCAKXKAKAKAKADARDKRRMDKDKRRAKAKA	773			
QY	636 AGGATCTGAGA-AATGCTTGGGAAGAAAGTGGGATCTCCAGATGATCAAGAAACATTGGA	694			
DB	774 AKAADDDADADAKADAKADAKADAKADADADGREGGKKRRKRAKRRKKKKKDDAMGDKKKAK	833			
QY	695 TAGAAGATTCATTTATAGAAAGATGTCCTCAAAAAGCAAAAAGACCAATTTCATAG	754			
DB	834 KDRKAAAKAKADADADAKAKRRRADDKRRKADAKAAKAKKKDDDDAKAATKAKKATKAK	893			
QY	755 GTGAGTTTGTGCTCTACAGAAAAGCTGACCTTGATCAAGATTAATAATGGACCTCCTTG	814			
DB	894 KDKAKAKKKKKKKKDKAKAKAKADADADKDDKDDKDDKDKADKKKWKDRKDRKDKAKK	953			
QY	815 TCGTTGGGAATGAGGAAAGAGATGAGAGCTGGCTACTAGTGGGAATTTTGTTCGGTT	874			
DB	954 DDDKKDKKDDKDDKDDADAKAMKAKKADADAKKAKAKAKAKAADDADAADAKAKAADAKA	1013			
QY	875 TTGCTGTTTATGATCTCTGAGAAAACCTGGAATTAAGATCTTTTGAAGACTTAAAG	934			
DB	1014 KAKKAKDDAAKAKKADAKADAKADAKAKAKRAKAKAKADADAKAKAKADKDDKDDAKDD	1073			
QY	935 GAGGAGATGACATTGA	950			

RESULT 6	CNS016E2/c	1204 bp	DNA	linear	GSS 26-JUL-1999
LOCUS	CNS016E2				
DEFINITION	Drosophila melanogaster genome survey sequence 17 end of BAC BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL106628				
VERSION	ALI06628.1	GI:5622852			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.				
FEATURES	location/Qualifiers				
source	1..1204 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACN15A12" /clone_11b="DrosBAC" /plasmid="pBelOBAC11" /note="end : 77"				
BASE COUNT	298 a 172 c 106 g 316 t 312 others				
ORIGIN					
Query Match	4.8%; Score 45.8; DB 17; Length 1204;				
Best Local Similarity	21.7%; Pred. No. 1.8; 119; Indels 0; Gaps 0;				
Matches	64; Conservative 112; Mismatches				
658	AGAACTGGGCGATCCACATGATGACGAACATTGATAGAGAATTGATTATAGAA 717				
1180	ATTAATKMKKAKKAKKAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 1121				
718	GAATGGTCAAAAAGAACAAAGACCAATTTCATAGTGGACTTGGTCCACAGAA 777				
1120	ADAAAAKAKKAAKAAAAAADAAAKKKRAMAAAAAAKAAKAAATKKDDDDKKRAAD 1061				
778	GCAGACCTGGAATCAAGCAATTAATGACACCTCTTGTGCTGCGGAATGAGAAAAG 837				
1060	AKKKKKKAKADAKADAKDKDKKAKADADDDADDAKKKKKKDKKKDDATDADMAAK 1001				
838	AGATGAGAGCTGGGCATACCTGGAATTGTTCGGGTTTGGTGTTAATGATCTCTGAGA 897				
1000	AAAATDKKAKKKKKDKKKKAAKAAKDKKKKKKKKKKKKKKKKKKAAADAAA 941				
898	AAACCTGGAAATTAAGATCTTTAGAGCTTAAATGAGAGAGATAGCATTTGAAT 952				
940	AAAAAATKKKKKADAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 886				

ACCESSION	fly), genomic survey sequence.					
VERSION	AL078812					
KEYWORDS	GSS.					
SOURCE	Drosophila melanogaster.					
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epiphytoidae; Drosophilidae; Drosophila. 1 (bases 1 to 1101)					
REFERENCE	Genoscope.					
AUTHORS	Direct Submission					
TITLE	Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :					
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr) - Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aron Mammosser at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.					
FEATURES						
Source	Location/Qualifiers 1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACR48119" /clone.lib="RPCI-98" /note="end : TET3"					
BASE COUNT	206 a 185 c 183 g 354 t 173 others					
ORIGIN						
Query Match	4.8%; Score 45.4; DB 17; Length 1101;					
Best Local Similarity	37.6%; Pred. No.2,2; Indels 3; Gaps 1;					
Matches	94; Conservative 47; Mismatches 106; Indels 3; Gaps 1;					
OY	708 TTTTATAGAGATGCTCAAAAAGAACAAGAACCATTTACATAGTGTAAGTTGGTCG 767					
Dd	999 KTWAAACGAAGAGCGCTTAARAAAAARGRAAANKACTAATTAGAKMAHKHKKMGATT 940					
OY	768 CTACAGAAAG---GTGACCTTGANTCAAAGATAAATGAGACCCTCTTGTGCCTCCGCA 824					
Dd	939 AKAMGBTRAAKTTNATTTAATTAKAMWTRAAMTAKAMRGACTGARGGGGARGGTGAGG 880					
OY	825 AATGAGAAAGAGATGGAGCTGGGCATATCTGGGAATTTTGTCCGTTTGGTGTGTTA 884					
Dd	879 GGRGRRGAAAGAGRARAAAAGDTRCAMWARAKKGCGGGRAMAGCAABAGWTGAAKTTTK 820					
OY	885 TGATCTCTGAGAAAACCTGGAATTAAGAATCTTTTAGAGCTTTAATATAGGAGGAGATAG 944					
Dd	819 WAAAGAKTWRTAKKAAMATGCGCAATADAAAATWATPAAKATATTAATAATTKAAASAMTAA 760					
OY	945 CATTGAATTA 954					
Dd	759 AKTAGAASAW 750					
RESULT 8						
CNS00CF2	1031 bp DNA linear GSS 04-JUN-1999					
LOCUS	Drosophila melanogaster genome survey sequence TET3 end of BAC #					
DEFINITION	BACR23K05 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.					
ACCESSION	AL059199					
VERSION	1 GI:4946662					

KEYWORDS	GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
	Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1031)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .
FEATURES	
source	1..1031 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACR25K05" /clone_1lb="RPc1-98" /note="end : TET3"
BASE COUNT	176 a 169 c 124 g 214 t 348 others
ORIGIN	
Query Match	4.7%, Score 44.8; DB 17; Length 1031;
Best Local Similarity	26.6%; Pred. No. 3.1;
Matches	71; Conservative 80; Mismatches 116; Indels 0; Gaps 0;
QY	619 GCAGCTGAGTGGGCGTGAAGCATCTAGAAATGCTTGGGAGAGAACTGGGCGATCTCCAGAT 678
Db	722 RGGGAGRGKRRKAGARRRKAGTGMGGGKAGGTWKKAAAMAKRTGGGWRKGRKAKRWW 781
QY	679 GATCGAACAATTCATGATGAGAATTCATTTATAGAGAAGATGGTCAAAAAGAACAA 738
Db	782 KRTTTCGRRGDTWAKGKTGRRAKKNMTWMAARAADADMDARADTTAGAGGDMWMAAAR 841
QY	739 AGACCAATTTACATAGTGTGAGTTGGTCCCTACAGAAAAGCTGACCTTGAATCAAGAATA 798
Db	842 RGRKAGATADRTAGAMRDKRRWDKDKAGTDRKTADTKRRADARADADRGAGTAMGAMA 901
QY	799 AAATGGACCTCTTGTGCTTCGCGCAATGGAGAAAAGAGATGGAGCTGGCGATACG 858
Db	902 WAMAWKAARTAGTGTDRKATRTADTAAGRNRAMWATRRKAWAGAGAKATWTTTDD 961
QY	859 GAATTTTGTCCGGTTTGGTGTTAT 885
Db	962 TWTCTTTGTNTKTKTTTWTWTTTITKT 988
RESULT 9	
CNS015YX/c	CNS015YX
LOCUS	915 bp DNA linear GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence Sfb end of BAC BACN15N15 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL106083
VERSION	AL106083.1 GI:5619993
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster



GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 09:27:47 ; Search time 46 seconds
(without alignments)
6360.209 Million cell updates/sec

Title: US-09-880-729A-1

Perfect score: 954
Sequence: 1 atgggtgttgatccttga.....gagagatagcatgataa 954

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	954	100.0	954	2	US-09-066-075-1
2	954	100.0	954	2	US-08-518-615A-1
3	954	100.0	954	3	US-08-951-889-1
4	954	100.0	954	4	US-09-472-857-1
5	235.8	24.7	1083	4	US-09-134-078-13
6	73	7.7	1826	4	US-09-286-691-11
7	73	7.7	1826	4	US-09-687-147-11
8	68.6	7.2	1532	3	US-09-118-324-1
9	62.8	6.6	1170	4	US-09-419-459-3
10	62.8	6.6	1182	4	US-09-419-459-9
11	62.8	6.6	2110	4	US-09-419-459-1
12	60.4	6.3	7218	4	US-08-232-463-14
13	49.4	5.2	1614	4	US-09-110-959A-3
14	48.2	5.1	2774	3	US-08-732-433-2
15	40	4.2	2930	1	US-08-356-354-5
16	40	4.2	2930	2	US-08-778-656-5
17	40	4.2	3625	2	US-08-356-354-3
18	40	4.2	3625	2	US-08-778-656-3
19	40	4.2	3740	1	US-08-356-354-1
20	40	4.2	3740	2	US-08-778-656-1
21	36.4	3.8	1299	1	US-08-463-262A-1
22	36.4	3.8	1299	1	US-08-463-989-1
23	36.4	3.8	1299	4	US-09-003-574-1
24	36.4	3.8	1299	4	US-09-003-570-1
25	36.4	3.8	2028	4	US-09-003-574-33
26	36.4	3.8	2028	4	US-09-003-570-33
27	36.4	3.8	2032	4	US-09-003-574-32

28	36.4	3.8	2032	4	US-09-003-570-32	Sequence 32, Appl
29	36.4	3.8	2076	4	US-09-003-574-30	Sequence 30, Appl
30	36.4	3.8	2076	4	US-09-003-570-30	Sequence 30, Appl
31	36.4	3.8	2126	1	US-08-463-262A-2	Sequence 2, Appl
32	36.4	3.8	2126	1	US-08-463-989-2	Sequence 2, Appl
33	36.4	3.8	2126	4	US-09-003-574-2	Sequence 2, Appl
34	36.4	3.8	2126	4	US-09-003-574-2	Sequence 2, Appl
35	36.4	3.8	2292	4	US-09-003-574-29	Sequence 29, Appl
36	36.4	3.8	2292	4	US-09-003-570-29	Sequence 29, Appl
37	35.4	3.7	2323	2	US-07-862-588B-3	Sequence 3, Appl
38	35	3.7	1074	3	US-09-248-335-67	Sequence 67, Appl
39	34.8	3.6	87350	3	US-08-781-891-79	Sequence 79, Appl
40	34.8	3.6	87543	4	US-09-791-211-3	Sequence 3, Appl
41	34.6	3.6	1816	4	US-09-149-476-167	Sequence 167, App
42	34.2	3.6	2107	4	US-09-180-852-1	Sequence 1, Appl
43	34.2	3.6	19124	2	US-08-487-826B-13	Sequence 13, Appl
44	34	3.6	2236	4	US-09-697-367-1	Sequence 1, Appl
45	34	3.6	4450	3	US-08-617-860B-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-09-066-075-1
; Sequence 1, Application US/09066075
; Patent No. 5925749
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: CARCELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,075
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615
; FILING DATE: August 23, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 954 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; US-09-066-075-1
;
Query Match          100.0%; Score 954; DB 2; Length 954;
Best Local Similarity 100.0%; Pred. No. 5.4e+259;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGGGTGTGATCCTTTTAAAGAACAAATTTTGGAGAGCATTTATATAGCAAT 60
|||||
```

```

Db      1 ATGGGTGTGATCCTTTTGAAGAACAAATATTTGGAGAGGCAATTAATAGCAAT 60
QY      61 GCGCTTGAAGCACCACCAATAGAGGAGCTGGGAGTGGTGTATTAAGATGTTCTTGAC 120
      |||||||
Db      61 GCGCTTGAAGCACCACCAATAGAGGAGCTGGGAGTGGTGTATTAAGATGTTCTTGAC 120
QY      121 ATTATTAAGAGGCGGTTCTCATGTTCGAATTCACATAGATGAGTACGACGCT 180
      |||||||
Db      121 ATTATTAAGAGGCGGTTCTCATGTTCGAATTCACATAGATGAGTACGACGCT 180
QY      121 ATTATTAAGAGGCGGTTCTCATGTTCGAATTCACATAGATGAGTACGACGCT 180
      |||||||
Db      121 ATTATTAAGAGGCGGTTCTCATGTTCGAATTCACATAGATGAGTACGACGCT 180
QY      181 TACGCTTCTCTCTTATTAATCATGTGCTTCTTCAAAAGAGTGTGATGAGTGATA 240
      |||||||
Db      181 TACGCTTCTCTCTTATTAATCATGTGCTTCTTCAAAAGAGTGTGATGAGTGATA 240
QY      181 TACGCTTCTCTCTTATTAATCATGTGCTTCTTCAAAAGAGTGTGATGAGTGATA 240
      |||||||
Db      241 AACGGAGCCCGAAAGAGAGCTGGCTGTCTATTAATATTCATCACTACAGAGATTA 300
      |||||||
Db      241 AACGGAGCCCGAAAGAGAGCTGGCTGTCTATTAATATTCATCACTACAGAGATTA 300
QY      301 ATGAATGATCCAGAGAACACAGAGAAATTTCTTCTCTTTGGAACCAATTCGTAT 360
      |||||||
Db      301 ATGAATGATCCAGAGAACACAGAGAAATTTCTTCTCTTTGGAACCAATTCGTAT 360
QY      301 ATGAATGATCCAGAGAACACAGAGAAATTTCTTCTCTTTGGAACCAATTCGTAT 360
      |||||||
Db      361 CGTTATTAAGACTATCCGGAACCTATTTTGAATTCGAATGAACCTCAGGAAAT 420
      |||||||
QY      361 CGTTATTAAGACTATCCGGAACCTATTTTGAATTCGAATGAACCTCAGGAAAT 420
      |||||||
Db      421 CTACTCCGGAATAATGGAATGAACTGCTTGAAGAGCTCTAAGATTATAGATCAAT 480
      |||||||
QY      421 CTACTCCGGAATAATGGAATGAACTGCTTGAAGAGCTCTAAGATTATAGATCAAT 480
      |||||||
Db      421 CTACTCCGGAATAATGGAATGAACTGCTTGAAGAGCTCTAAGATTATAGATCAAT 480
QY      481 GACCAAAAGACACATATATTTATAGCAGACGTGAATGGGGGGTATCTCCCTTGA 540
      |||||||
Db      481 GACCAAAAGACACATATATTTATAGCAGACGTGAATGGGGGGTATCTCCCTTGA 540
QY      481 GACCAAAAGACACATATATTTATAGCAGACGTGAATGGGGGGTATCTCCCTTGA 540
      |||||||
Db      541 AAACGTCTGTCTCCAAATGGAATAATTTCTATAGTCAATTCACACTACATCAATCT 600
      |||||||
QY      541 AAACGTCTGTCTCCAAATGGAATAATTTCTATAGTCAATTCACACTACATCAATCT 600
      |||||||
Db      601 TTGCAATTTACCCATCAAGAGCTGAGTGGTGGAGGAGTGAAGAAATGGTGGGAGA 660
      |||||||
QY      601 TTGCAATTTACCCATCAAGAGCTGAGTGGTGGAGGAGTGAAGAAATGGTGGGAGA 660
      |||||||
Db      661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTATAGAGAA 720
      |||||||
QY      661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTATAGAGAA 720
      |||||||
Db      721 TGCTCAAAAAGAACAAAGACCAATTTACATAGCTGAGTGGTGGCTACAGAAAAGCT 780
      |||||||
QY      721 TGCTCAAAAAGAACAAAGACCAATTTACATAGCTGAGTGGTGGCTACAGAAAAGCT 780
      |||||||
Db      721 TGCTCAAAAAGAACAAAGACCAATTTACATAGCTGAGTGGTGGCTACAGAAAAGCT 780
QY      781 GACCTGAATCAAGATTAATGAGACCTCTTGTCTGCGCAATGAGAAAAGAGAGA 840
      |||||||
Db      781 GACCTGAATCAAGATTAATGAGACCTCTTGTCTGCGCAATGAGAAAAGAGAGA 840
QY      841 TGGAGCTGGGCATAGTGGAAATTTGTCGCGTTTGTGTTATGATCTGAGAGAA 900
      |||||||
Db      841 TGGAGCTGGGCATAGTGGAAATTTGTCGCGTTTGTGTTATGATCTGAGAGAA 900
QY      841 TGGAGCTGGGCATAGTGGAAATTTGTCGCGTTTGTGTTATGATCTGAGAGAA 900
      |||||||
Db      901 ACCTGGAATTAAGATCTTTTAGAAGCTTTATAGAGAGAGATGACATTGATA 954
      |||||||
QY      901 ACCTGGAATTAAGATCTTTTAGAAGCTTTATAGAGAGAGATGACATTGATA 954
      |||||||
Db      901 ACCTGGAATTAAGATCTTTTAGAAGCTTTATAGAGAGAGATGACATTGATA 954

```

RESULT 2

```

US-08-518-615A-1
; Sequence 1, Application US/08518615A
; Patent No. 5962238
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

```

```

; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615A
; FILING DATE: August 23, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 954 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
;
US-08-518-615A-1
;
Query Match 100.0%; Score 954; DB 2; Length 954;
Best Local Similarity 100.0%; Pred. No. 5,4e-259;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY      1 ATGGGTGTGATCCTTTTGAAGAACAAATATTTGGAGAGGCAATTAATAGCAAT 60
      |||||||
Db      1 ATGGGTGTGATCCTTTTGAAGAACAAATATTTGGAGAGGCAATTAATAGCAAT 60
;
QY      61 GCGCTTGAAGCACCACCAATAGAGGAGCTGGGAGTGGTGTATTAAGATGTTCTTGAC 120
      |||||||
Db      61 GCGCTTGAAGCACCACCAATAGAGGAGCTGGGAGTGGTGTATTAAGATGTTCTTGAC 120
;
QY      121 ATTATTAAGAGGCGGTTCTCATGTTCGAATTCACATAGATGAGTACGACGCT 180
      |||||||
Db      121 ATTATTAAGAGGCGGTTCTCATGTTCGAATTCACATAGATGAGTACGACGCT 180
;
QY      181 TACGCTTCTCTCTTATTAATCATGTGCTTCTTCAAAAGAGTGTGATGAGTGATA 240
      |||||||
Db      181 TACGCTTCTCTCTTATTAATCATGTGCTTCTTCAAAAGAGTGTGATGAGTGATA 240
;
QY      241 AACGGAGCCCGAAAGAGAGCTGGCTGTCTATTAATATTCATCACTACAGAGATTA 300
      |||||||
Db      241 AACGGAGCCCGAAAGAGAGCTGGCTGTCTATTAATATTCATCACTACAGAGATTA 300
;
QY      301 ATGAATGATCCAGAGAACACAGAGAAATTTCTTCTCTTTGGAACCAATTCGTAT 360
      |||||||
Db      301 ATGAATGATCCAGAGAACACAGAGAAATTTCTTCTCTTTGGAACCAATTCGTAT 360
;
QY      361 CGTTATTAAGACTATCCGGAACCTATTTTGAATTCGAATGAACCTCAGGAAAT 420
      |||||||
Db      361 CGTTATTAAGACTATCCGGAACCTATTTTGAATTCGAATGAACCTCAGGAAAT 420
;
QY      421 CTACTCCGGAATAATGGAATGAACTGCTTGAAGAGCTCTAAGATTATAGATCAAT 480
      |||||||
Db      421 CTACTCCGGAATAATGGAATGAACTGCTTGAAGAGCTCTAAGATTATAGATCAAT 480
;
QY      481 GACCAAAAGACACATATATTTATAGCAGACGTGAATGGGGGGTATCTCCCTTGA 540
      |||||||
Db      481 GACCAAAAGACACATATATTTATAGCAGACGTGAATGGGGGGTATCTCCCTTGA 540

```

QY 541 AACTGTCTGTCCCAAAATGGAAAAATTCATAGTACAAATTCACATCT 600
 Db 541 AACTGTCTGTCCCAAAATGGAAAAATTCATAGTACAAATTCACATCT 600
 QY 601 TTGCAATTTACCATCAAGAGCTGAGTGGTGGAAGAGCTGAGAAATGTTGGGAGA 660
 Db 601 TTGCAATTTACCATCAAGAGCTGAGTGGTGGAAGAGCTGAGAAATGTTGGGAGA 660
 QY 661 AAGGGGATCTCCAGATGATGAGAAATTCATAGTACAAATTCACATCT 720
 Db 661 AAGGGGATCTCCAGATGATGAGAAATTCATAGTACAAATTCACATCT 720
 QY 721 TGGTCAAAAAAGAACAAAGACCAATTTACATAGTATGTTGGTCTCAGAAAAGCT 780
 Db 721 TGGTCAAAAAAGAACAAAGACCAATTTACATAGTATGTTGGTCTCAGAAAAGCT 780
 QY 781 GACCTGATCAAGATTAATTAATGACCTCTTGTGCTGCGAAATGAGAAAAAGAGA 840
 Db 781 GACCTGATCAAGATTAATTAATGACCTCTTGTGCTGCGAAATGAGAAAAAGAGA 840
 QY 841 TGGAGCTGGGCATACCTGGGAATTTGTTCCGGTTTGGTGTATGATCTGAGAAA 900
 Db 841 TGGAGCTGGGCATACCTGGGAATTTGTTCCGGTTTGGTGTATGATCTGAGAAA 900
 QY 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAAATAGAGAGATAGCATTTGATTA 954
 Db 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAAATAGAGAGATAGCATTTGATTA 954

RESULT 3

US-08-951-889-1
 ; Sequence 1, Application US/08951889
 ; Patent No. 6008032

GENERAL INFORMATION:

APPLICANT: Mathur, E., et al.
 TITLE OF INVENTION: Carboxymethyl Cellulase from
 TITLE OF INVENTION: Thermotoga Maritima
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
 STREET: 6 BECKER FARM ROAD
 CITY: ROSELAND
 STATE: NEW JERSEY
 COUNTRY: USA
 ZIP: 07068
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 INCH DISKETTE
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WORD PERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/951,889
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/518,615
 FILING DATE: August 23, 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: FERRARO, GREGORY D.
 REGISTRATION NUMBER: 36,134
 REFERENCE/DOCKET NUMBER: 331400-20
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-11700
 TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
 LENGTH: 954 BASE PAIRS
 TYPE: NUCLEIC ACID
 STRANDEDNESS: SINGLE
 TOPOLOGY: LINEAR
 MOLECULE TYPE: CDNA

US-08-951-889-1

Query Match 100.0%; Score 954; DB 3; Length 954;
 Best Local Similarity 100.0%; Pred. No. 5.4e-259;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTGTGATCCCTTTTGAAGAGAACAAATATTTGGGAGAGCATTAATATAGAAAT 60
 Db 1 ATGGGTGTGATCCCTTTTGAAGAGAACAAATATTTGGGAGAGCATTAATATAGAAAT 60
 QY 61 GCGCTTAAGACCAATATGAGGAGACTGGGAGTGTATTAAGATGAGTCTTGAC 120
 Db 61 GCGCTTAAGACCAATATGAGGAGACTGGGAGTGTATTAAGATGAGTCTTGAC 120
 QY 121 ATTATTAAGAGAGCGGTTCTCTCATGTGCAATTCATTAAGATGAGTACAGCAGCT 180
 Db 121 ATTATTAAGAGAGCGGTTCTCTCATGTGCAATTCATTAAGATGAGTACAGCAGCT 180
 QY 181 TACGCTTTCTCCTTTTAAATCATGATCGTCTTCAAAAAGATGATGATGATA 240
 Db 181 TACGCTTTCTCCTTTTAAATCATGATCGTCTTCAAAAAGATGATGATGATA 240
 QY 241 AACGAGCCCTGAAAGAGAGAGCTGCTGTTGCTATTAATATTCATCTACAGAGTTA 300
 Db 241 AACGAGCCCTGAAAGAGAGAGCTGCTGTTGCTATTAATATTCATCTACAGAGTTA 300
 QY 301 ATGAATGATCCAGAGAAGACCAAGAAAGATTTCTGCTCTTGGAAACAAATGCTGAT 360
 Db 301 ATGAATGATCCAGAGAAGACCAAGAAAGATTTCTGCTCTTGGAAACAAATGCTGAT 360
 QY 361 CGTTATTAAGACTATCCCGAAACTCTATTTTGAATTTCTGAATGAACCTCAGGAAAT 420
 Db 361 CGTTATTAAGACTATCCCGAAACTCTATTTTGAATTTCTGAATGAACCTCAGGAAAT 420
 QY 421 CTTACTCCGGAATAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 Db 421 CTTACTCCGGAATAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 QY 481 GACAAAAGACACTATTAATTAATGAGCAGACGTAATGGGGGGTATATCGCCCTGAA 540
 Db 481 GACAAAAGACACTATTAATTAATGAGCAGACGTAATGGGGGGTATATCTGCCCTGAA 540
 QY 541 AACTGTCTGTCCCAAAATGGAAAAATTCATAGTACAAATTCACATCT 600
 Db 541 AACTGTCTGTCCCAAAATGGAAAAATTCATAGTACAAATTCACATCT 600
 QY 601 TTGCAATTTACCATCAAGAGCTGAGTGGTGGAAGAGCTGAGAAATGTTGGGAGA 660
 Db 601 TTGCAATTTACCATCAAGAGCTGAGTGGTGGAAGAGCTGAGAAATGTTGGGAGA 660
 QY 661 AAGGGGATCTCCAGATGATGAGAAATTCATAGTACAAATTCACATCT 720
 Db 661 AAGGGGATCTCCAGATGATGAGAAATTCATAGTACAAATTCACATCT 720
 QY 721 TGGTCAAAAAAGAACAAAGACCAATTTACATAGTATGTTGGTCTCAGAAAAGCT 780
 Db 721 TGGTCAAAAAAGAACAAAGACCAATTTACATAGTATGTTGGTCTCAGAAAAGCT 780
 QY 781 GACCTGATCAAGATTAATTAATGACCTCTTGTGCTGCGAAATGAGAAAAAGAGA 840
 Db 781 GACCTGATCAAGATTAATTAATGACCTCTTGTGCTGCGAAATGAGAAAAAGAGA 840
 QY 841 TGGAGCTGGGCATACCTGGGAATTTGTTCCGGTTTGGTGTATGATCTGAGAAA 900
 Db 841 TGGAGCTGGGCATACCTGGGAATTTGTTCCGGTTTGGTGTATGATCTGAGAAA 900
 QY 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAAATAGAGAGATAGCATTTGATTA 954
 Db 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAAATAGAGAGATAGCATTTGATTA 954

RESULT 4

US-09-472-857-1

Sequence 1, Application US/09472857
Patent No. 6245547
GENERAL INFORMATION:
APPLICANT: Mathur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from
TITLE OF INVENTION: Thermotoga Maritima
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/472,857
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/951,889
FILING DATE:
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-09-472-857-1

Query Match 100.0%; Score 954; DB 4; Length 954;
Best Local Similarity 100.0%; Pred. No. 5,4e-259;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTTGATCGCTTTTGAAGAACAAATATTGGAGAGCGATTATATAGAAAT 60
DB 1 ATGGGTTGATCGCTTTTGAAGAACAAATATTGGAGAGCGATTATATAGAAAT 60
QY 61 GCGCTTGAAACACCAATGAGGAGACTGGGAGTGTGATATAAAGATGAGTTCTTGAC 120
DB 61 GCGCTTGAAACACCAATGAGGAGACTGGGAGTGTGATATAAAGATGAGTTCTTGAC 120
QY 121 ATTATATAAAGAGCGGTTTCTCTCATGTCGATTCGATTAAGATGAGAGACCGCT 180
DB 121 ATTATATAAAGAGCGGTTTCTCTCATGTCGATTCGATTAAGATGAGAGACCGCT 180
QY 181 TAGCGCTTCTCTCTTATAAATCATGATCGCTTCTTCAAAAGAGTGATGAAGTATA 240
DB 181 TAGCGCTTCTCTCTTATAAATCATGATCGCTTCTTCAAAAGAGTGATGAAGTATA 240
QY 241 AACGAGACCTTGAAAGAGAGCTGGCTGTTCTTAATATATTCATCTACGAGAGATA 300
DB 241 AACGAGACCTTGAAAGAGAGCTGGCTGTTCTTAATATATTCATCTACGAGAGATA 300
QY 301 ATGAATGATCCAGAAAGACACAGAAAGATTCTTGCTCTTTGAAACAAATTCGTGAT 360
DB 301 ATGAATGATCCAGAAAGACACAGAAAGATTCTTGCTCTTTGAAACAAATTCGTGAT 360

QY 361 CATTATTAAGACTATCCCGAAACCTCTATTTTGAATTCGATGAGACCTCAGCAAT 420
DB 361 CATTATTAAGACTATCCCGAAACCTCTATTTTGAATTCGATGAGACCTCAGCAAT 420
QY 421 CTCTACCGGAAAGATGAGTGAAGTCTTGAGAGACCTCTAAAGTTATAGATCAAT 480
DB 421 CTCTACCGGAAAGATGAGTGAAGTCTTGAGAGACCTCTAAAGTTATAGATCAAT 480
QY 481 GACAAAGACACCTATATATATAGCAGACCTGAATGGGGGGGTATATGCCCTTGA 540
DB 481 GACAAAGACACCTATATATATAGCAGACCTGAATGGGGGGGTATATGCCCTTGA 540
QY 541 AAAGTGTCTGCCCAAAATGSGAAAAAATCTATAGTCAATTCACCTACTCAATCT 600
DB 541 AAAGTGTCTGCCCAAAATGSGAAAAAATCTATAGTCAATTCACCTACTCAATCT 600
QY 601 TTGCAATTTACCATCAAGAGAGCTGAGTGGTGAAGATCTGAGAAATGGTTGGAGA 660
DB 601 TTGCAATTTACCATCAAGAGAGCTGAGTGGTGAAGATCTGAGAAATGGTTGGAGA 660
QY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTATAGAA 720
DB 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTATAGAA 720
QY 721 TGCTCAAAAGACAAAGACCAATTTACATAGTGAATTTGCTGCTACAGAAAGCT 780
DB 721 TGCTCAAAAGACAAAGACCAATTTACATAGTGAATTTGCTGCTACAGAAAGCT 780
QY 781 GACCTTGAATCAAGATTAATAGACCTCTTTGCTGCTGCGGAAATGGAGAAAGAGA 840
DB 781 GACCTTGAATCAAGATTAATAGACCTCTTTGCTGCTGCGGAAATGGAGAAAGAGA 840
QY 841 TGGAGCTGGGACTACTGGGAATTTGTCGCTTTGCTGTTATGATCTCTGAGAAA 900
DB 841 TGGAGCTGGGACTACTGGGAATTTGTCGCTTTGCTGTTATGATCTCTGAGAAA 900
QY 901 ACCTGAATTAAGATCTTTTGAAGCTTTAATAGAGAGATAGATGATTA 954
DB 901 ACCTGAATTAAGATCTTTTGAAGCTTTAATAGAGAGATAGATGATTA 954

RESULT 5
US-09-134-078-13
Sequence 13, Application US/09134078
Patent No. 6368844
GENERAL INFORMATION:
APPLICANT: Byline, Edward J.
TITLE OF INVENTION: GLYCOSIDASE ENZYMES
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,078
FILING DATE: 13-AUG-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/949,026
FILING DATE: 10-OCT-1997
APPLICATION NUMBER: 60/056,916
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347

Best Local Similarity 49.7%; Pred. No. 2.4e-08;
Matches 194; Conservative 0; Mismatches 187; Indels 9; Gaps 1;

```
QY 99 GATTAAGATGAGTCTTCTGCACATTTATAAAGAGCGGTTCTCTCATGTTGGAATTC 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 GACTACTGAAGATATGTTCAGAGTTTAAACGACCAACCAATTAATGCTTCGTAATTC 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 AATAAGATGAGTACGACGCTTACGCTTCTCTTATTAATAATCATGATGCTTCTT 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 288 AACTACTGTGCTGCTACCTCGTGAAGCTCCAGATTATTAAGATTGATGAATAATGTT 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 219 CAAAGAGTGGATGAAGTGTAAACGAGCCGTAAAGAGAGAGCTGGCTGTCTATAAA 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 348 AAGAGAGTTCATGAAGTGTGTATTCATACAGAGAGAGCATTTGTATCTTAA 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 279 TATTCACTACAG-----AGAGTTAATGATGATCCAGAAACACAGAAAG 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 TCTTCATCATGAACCTGGAATCATGCTTCTGTGAACCTTGTGAACAGCAAGAAAT 467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 330 ATTCTTGTCTTGGAAACAAATGCTGATCGTTAATAAGACTATCCGAACCTTAT 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 TTAGAGAAAGATCTGCTCAATTCGTGAAGATTTAAGATTATGATGAACACTTAAT 527
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 390 TTTTGAATTTCTGAATGAACCTCAGGAAATCTTACTCCGAAAAATGGAATGACTGCT 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 528 CTTCGAAGATTAACGAAACCAAGAAAGATGATACTCCAGTTGAATGAGCTGTGTGA 587
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 450 TGAGGAAGCTCTAAAGTTAATGATCAAT 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 588 TCAAGAAAGTTGGATGCTGTATATGCTAT 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 10

```
US-09-419-459-9
; Sequence 9, Application US/09419459
; Patent No. 6222028
; GENERAL INFORMATION:
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Cheng, Kuo-Joan
; APPLICANT: Tsai, Cheng-Pang
; APPLICANT: Chang, Chia-Chieh
; TITLE OF INVENTION: CELLULOSE ENZYMES
; FILE REFERENCE: 08919/037001
; CURRENT APPLICATION NUMBER: US/09/419,459
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Pityomyces rhizinflata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1182)
US-09-419-459-9
```

Query Match
Best Local Similarity 49.7%; Pred. No. 2.4e-08;
Matches 194; Conservative 0; Mismatches 187; Indels 9; Gaps 1;

```
QY 99 GATTAAGATGAGTCTTCTGCACATTTATAAAGAGCGGTTCTCTCATGTTGGAATTC 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 GACTACTGAAGATATGTTCAGAGTTTAAACGACCAACCAATTAATGCTTCGTAATTC 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 AATAAGATGAGTACGACGCTTACGCTTCTCTTATTAATAATCATGATGCTTCTT 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 AACTACTGTGCTGCTACCTCGTGAAGCTCCAGATTATTAAGATTGATGAATAATGTT 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 219 CAAAGAGTGGATGAAGTGTAAACGAGCCGTAAAGAGAGAGCTGGCTGTCTATAAA 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 AAGAGAGTTCATGAAGTGTGTATTCATACAGAGAGAGCATTTGTATCTTAA 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 279 TATTCACTACAG-----AGAGTTAATGATGATCCAGAAACACAGAAAG 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 420 TCTTCATCATGAACCTGGAATCATGCTTCTGTGAACCTTGTGAACAGCAAGAAAT 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 330 ATTCTTGTCTTGGAAACAAATGCTGATCGTTAATAAGACTATCCGAACCTTAT 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 TTAGAGAAAGATCTGCTCAATTCGTGAAGATTTAAGATTATGATGAACACTTAAT 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 390 TTTTGAATTTCTGAATGAACCTCAGGAAATCTTACTCCGAAAAATGGAATGACTGCT 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 540 CTTCGAAGATTAACGAAACCAAGAAAGATGATACTCCAGTTGAATGAGCTGTGTGA 599
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 450 TGAGGAAGCTCTAAAGTTAATGATCAAT 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 600 TCAAGAAAGTTGGATGCTGTATATGCTAT 629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 11

```
US-09-419-459-1
; Sequence 1, Application US/09419459
; Patent No. 6222028
; GENERAL INFORMATION:
; APPLICANT: Liu, Jin-Hao
; APPLICANT: Cheng, Kuo-Joan
; APPLICANT: Tsai, Cheng-Pang
; APPLICANT: Chang, Chia-Chieh
; TITLE OF INVENTION: CELLULOSE ENZYMES
; FILE REFERENCE: 08919/037001
; CURRENT APPLICATION NUMBER: US/09/419,459
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2110
; TYPE: DNA
; ORGANISM: Pityomyces rhizinflata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1749)
US-09-419-459-1
```

Query Match
Best Local Similarity 49.7%; Score 62.8; DB 4; Length 2110;
Matches 194; Conservative 0; Mismatches 187; Indels 9; Gaps 1;

```
QY 99 GATTAAGATGAGTCTTCTGCACATTTATAAAGAGCGGTTCTCTCATGTTGGAATTC 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 557 GACTACTGAAGATATGTTCAGAGTTTAAACGACCAACCAATTAATGCTTCGTAATTC 616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 AATAAGATGAGTACGACGCTTACGCTTCTCTTATTAATAATCATGATGCTTCTT 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 AACTACTGTGCTGCTACCTCGTGAAGCTCCAGATTATTAAGATTATGATGAATAATGTT 676
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 219 CAAAGAGTGGATGAAGTGTAAACGAGCCGTAAAGAGAGAGCTGGCTGTCTATAAA 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 677 AAGAGAGTTCATGAAGTGTGTATTCATACAGAGAGAGCATTTGTATCTTAA 736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 279 TATTCACTACAG-----AGAGTTAATGATGATCCAGAAACACAGAAAG 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 737 TCTTCATCATGAACCTGGAATCATGCTTCTGTGAACCTTGTGAACAGCAAGAAAT 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 330 ATTCTTGTCTTGGAAACAAATGCTGATCGTTAATAAGACTATCCGAACCTTAT 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 797 TTAGAGAAAGATCTGCTCAATTCGTGAAGATTTAAGATTATGATGAACACTTAAT 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 390 TTTTGAATTTCTGAATGAACCTCAGGAAATCTTACTCCGAAAAATGGAATGACTGCT 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 857 CTTCGAAGATTAACGAAACCAAGAAAGATGATACTCCAGTTGAATGAGCTGTGTGA 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 450 TGAGGAAGCTCTAAAGTTAATGATCAAT 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 917 TCAAGAAAGTTGGATGCTGTATATGCTAT 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 12
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHUEFLINGER, F.
APPLICANT: FALKNER, F.G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300. 6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pt-Fls
US-08-232-463-14

Query Match
Best Local Similarity 7.3%; Score 60.4; DB 1; Length 7218;
Matches 31; Conservative 222; Mismatches 173; Indels 0; Gaps 0;

QY 429 GGAATAATGGAATGCTGTTGAGGAGCTAAAGTTAATGAATTCATTCAGCAAAA 488
DB 1491 GCATCTGTAATTAATCTATCTATGCAAGTAGTTAAAGAGATAGAAATTTGGTGNRR 1432
QY 489 GCACATATATATATAGCAGAGCTGAGTGGGGGTATATCTGCCCTGAAAAACTGTC 548
DB 1431 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1372
QY 549 TGTCCCAAATGGGAAAAATTTCTATAGTTACATTCACCTACACAAATCCTTGCAT 608
DB 1371 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1312
QY 609 TAACCATCAGAGAGCTGAGTGGTGAAGATCGGAAATGTTGGAGAAAGTGGG 668
DB 1311 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1252
QY 669 ATCTCAGATGATCAGAAATTTGATGAGAAATTCATTTTATAGACATGCTCAA 728
DB 1251 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1192
```

```
QY 729 AAGAGCAAAAGACCAATTACATAGTGAGTTGCTGCTACAGAAAGCTGACTTCA 788
DB 1191 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1132
QY 789 ATCAAGATTAATGACCTCTTTCGTTCCGGAATGAGAAAGAGATGAGCTG 848
DB 1131 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1072
QY 849 GCACAT 854
DB 1071 RRRRRR 1066

RESULT 13
US-09-110-959A-3
; Sequence 3, Application US/09110959A
; Patent No. 6268197
GENERAL INFORMATION:
APPLICANT: Schuelein, Martin
APPLICANT: Outtrup, Helle
APPLICANT: Jorgensen, Per Lina
APPLICANT: Bjornvad, Mads Eskelund
TITLE OF INVENTION: Alkaline Xyloglucanase
FILE REFERENCE: 5206.200-US
CURRENT APPLICATION NUMBER: US/09/110,959A
CURRENT FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 0822/97
PRIOR FILING DATE: 1997-07-07
PRIOR APPLICATION NUMBER: 1213/97
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/054,039
PRIOR FILING DATE: 1997-07-28
PRIOR APPLICATION NUMBER: 60/063,694
PRIOR FILING DATE: 1997-10-28
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 3
LENGTH: 1614
TYPE: DNA
ORGANISM: Bacillus agaradhaerens NCIMB 40482
US-09-110-959A-3

Query Match
Best Local Similarity 5.2%; Score 49.4; DB 4; Length 1614;
Matches 186; Conservative 0; Mismatches 186; Indels 15; Gaps 1;

QY 101 TAAAGATGATGCTTTCGACATTAATAAGAAAGCCGTTTCTCTCATGCTTGAATGCA 160
DB 125 TAACAAGAGAGTTAATAAAMACGATTGCTGATGAGAGGTTAATAAGCATTCGTATCCAG 184
QY 161 TAAGATGAGAGTACGACGCTTACGGGTTTCTCTTATTAATCATGATGCTTCTTCA 220
DB 185 TGCAATGCAAAATCAATAGGTGTTCTCCAGATTATACGATTAATGAAATATATATCA 244
QY 221 AAGAGCTGATGAGATGAATGAAGGAGCCCTGAAAGAGAGCTGCTTCTTAATA 280
DB 245 ATCGGATGAGCAAGAGATGATGAGGCTTGGAGAGAGACTATATATGATTAATG 304
QY 281 TTATATACAGAGAGGTTAATGAATGATCCGAGAAACACA-----GG 325
DB 305 TGCAATGATGATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 364
QY 326 AAGATTTCTGCTCTTGGAAACAATTTGCTGATGATGATGATGATGATGATGATGATG 385
DB 365 CAAGATATACAGCTATTTGGAAATTTGCGAAAAATTCAAAGCCACATCCATAAGT 424
QY 386 TATTTTGAATTTGGAATGAACCTCAGAGAAATTTACTCCGAAAAATGATGATGAC 445
DB 425 TGATGTTGAGAGTGCATAGAGCTTACGAGAGAGTGGAGAGATTCAGAAA 484
QY 446 TGCTTGAGAGAGCTTAAGATTAAT 472
DB 485 ATCATCATGCTTACTTAGAATTTAA 511
```

```
RESULT 14
US-08-732-433-2
; Sequence 2, Application US/08732433
; Patent No. 6063611
; GENERAL INFORMATION:
; APPLICANT: van Solingen, Dieter
; TITLE OF INVENTION: No. 6063611el Alkaline Cellulase and
; TITLE OF INVENTION: Method of Producing Same
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/732,433
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05651
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC287-US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2774 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-732-433-2
Query Match 5.1%; Score 48.2; DB 3; Length 2774;
Best Local Similarity 50.2%; Pred. No. 0.00043;
Matches 159; Conservative 0; Mismatches 143; Indels 15; Gaps 1;

QY 109 GAGTCTTCGACATTAATAAAGAGCGGTTCTTCATGTTTCGAATTCGAATTAAGATGG 168
    || || || || || || || || || || || || || || || || || || || || ||
DB 835 GAATTAATGTAACGAGATTGGCGGATGAAGGGATATAAAGCATTCGGATTCGGTGACGTGG 894

QY 169 AGRACGACGCTTACGGCTTCCCTTATTAATCATGATGCGCTTCTTCAAAAGAGTG 228
    || || || || || || || || || || || || || || || || || || || || ||
DB 895 GAAATGTATCGAGGGGACCGGATTAATCATGATCCCGGTTTAAATCGAGTG 954

QY 229 GATGATGATTAACGAGCGCTGAAAGAGAGACTGCTGCTGATTAATAATATCATCAC 288
    || || || || || || || || || || || || || || || || || || || || ||
DB 955 GACCAATTTGTTCAATGGCGCTGGAAGAGATTGTATGATCATGATTAATTTACACCAT 1014

QY 289 TACGAGAGTTAATGAATGATCCAGAAACACAA-----GGAAGATTT 333
    || || || || || || || || || || || || || || || || || || || || ||
DB 1015 GATTCATGTTATGATGATTTATGAATGAGCACAACTACAACGCTGATGAGCCCAATAT 1074

QY 334 CTTCGCTTTGGAAACAATTTGCTGATCGTTATTAACATATCCGAAACTCTATTTT 393
    || || || || || || || || || || || || || || || || || || || || ||
DB 1075 CGCTCGCTGGGAGCAACTGAAACCACTTCAMAGACTATCCAAAGCTATATGTTT 1134

QY 394 GAAATCTGAATGAAC 410
    || || || || || || || || || || || || || || || || || || || || ||
DB 1135 GAAAGTCTCAATGAGCC 1151
```

```
RESULT 15
US-08-356-354-5
; Sequence 5, Application US/08356354
; Patent No. 5767365
; GENERAL INFORMATION:
; APPLICANT: SONNEWALD, Uwe
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
; TITLE OF INVENTION: PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostroienk, Faber, Geib & Sofien
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,354
; FILING DATE: 20-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/EP93/01605
; FILING DATE: 22-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P42 20 758.4
; FILING DATE: 24-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mellman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2930 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Solanum tuberosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 118..2841
; OTHER INFORMATION: /note= "Sucrose-Phosphatase-Synthase"
;
US-08-356-354-5
Query Match 4.2%; Score 40; DB 1; Length 2930;
Best Local Similarity 55.9%; Pred. No. 0.089;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 225 AGTGGATGAAGTGAACGAGCGCTGAAAGAGAGACTGCTGCTATTAATATCA 284
    || || || || || || || || || || || || || || || || || || || || ||
DB 1440 AGCTTATGAGTCTCCCAATGATGATCCAGCAAAAATGAGAGACTGTGATATACATAGGCT 1499

QY 285 TCACCTAGAGAGATTATGATGATCCAGAAAGACAGAGAAAGATTTCTGCTTTG 344
    || || || || || || || || || || || || || || || || || || || || ||
DB 1500 TCTTGACAAATGCTCTTACTGATCCCATGATCAGCAGGCAATTCCTGATGCTCTTT 1559

QY 345 GAAACAATTCGAT 360
    || || || || || || || || || || || || || || || || || || || || ||
DB 1560 GAAGTTGCTGCTGAT 1575
```

Search completed: April 26, 2003, 10:48:57
Job time: 61 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 10:19:02 ; Search time 99 Seconds
(without alignments)
10485.604 Million cell updates/sec

Title: US-09-880-729a-1

Perfect score: 954

Sequence: 1 atgagtgatgacatttga.....gagagatagcatgataca 954

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	954	100.0	954	9	US-09-880-729-1
2	235.8	24.7	1083	9	US-10-121-032-13
3	40	4.2	2930	10	US-09-376-045-5
4	40	4.2	3625	10	US-09-376-045-3
5	40	4.2	3740	10	US-09-376-045-1
6	37.8	4.0	1092	9	US-09-797-044-10
7	37.6	3.9	11474	10	US-09-816-028A-1
8	36.8	3.9	2000	9	US-09-938-842A-4603
9	36.8	3.9	3378	10	US-09-917-800A-1443
10	36.6	3.8	2392	9	US-09-797-464A-1
11	36.4	3.8	455	9	US-09-918-995-15849
12	36.4	3.8	1168	10	US-09-822-849A-492
13	36.4	3.8	1332	10	US-09-974-300-808
14	36.4	3.8	2447	9	US-10-091-504-2093
15	36.4	3.8	2639	9	US-10-091-504-2093
16	36.4	3.8	2639	10	US-09-764-869-2092
17	36.4	3.8	2640	9	US-09-764-869-2092
18	36.4	3.8	2640	10	US-10-091-504-2091
19	36.4	3.8	2640	10	US-09-764-869-2091

20	35.8	3.8	419	9	US-10-046-935-2068	Sequence 2068, Ap
21	35.8	3.8	419	9	US-09-878-178-2068	Sequence 2068, Ap
22	35.8	3.8	419	9	US-10-146-502-2068	Sequence 2068, Ap
23	35.6	3.7	640681	10	US-09-790-988-1	Sequence 1, Appli
24	35.4	3.7	6399	10	US-09-070-927A-76	Sequence 76, Appl
25	35.2	3.7	1149	10	US-09-880-107-3374	Sequence 3374, Ap
26	35	3.7	1431	10	US-09-822-849A-517	Sequence 517, App
27	35	3.7	2204	12	US-10-044-090-768	Sequence 768, App
28	34.8	3.6	9575	9	US-10-091-504-2412	Sequence 2412, Ap
29	34.8	3.6	9575	10	US-09-764-869-2412	Sequence 2412, Ap
30	34.6	3.6	1816	9	US-09-809-391-167	Sequence 167, App
31	34.6	3.6	368004	10	US-09-949-654-3	Sequence 167, App
32	34.4	3.6	927	10	US-09-834-975-750	Sequence 750, App
33	34.4	3.6	927	10	US-09-834-975-751	Sequence 751, App
34	34.2	3.6	2000	9	US-09-938-842A-3728	Sequence 3728, Ap
35	34.2	3.6	2150	9	US-09-981-353-45	Sequence 45, Appl
36	34	3.6	477	9	US-09-918-995-966	Sequence 966, App
37	34	3.6	2041	10	US-09-815-242-4352	Sequence 4352, Ap
38	34	3.6	2193	10	US-09-815-242-8391	Sequence 8391, Ap
39	34	3.6	2236	10	US-09-918-909-1	Sequence 1, Appli
40	34	3.6	84539	10	US-09-962-436-36	Sequence 36, Appl
41	33.8	3.5	2000	9	US-09-938-842A-3267	Sequence 3267, Ap
42	33.8	3.5	2004	10	US-09-887-576-211	Sequence 211, App
43	33.8	3.5	2004	10	US-09-887-576-297	Sequence 297, App
44	33.8	3.5	2746	7	US-08-781-986A-326	Sequence 326, App
45	33.8	3.5	6150	9	US-10-091-504-2311	Sequence 2311, Ap

ALIGNMENTS

RESULT 1
US-09-880-729-1
Sequence 1, Application US/09880729
Publication No. US2003004956A1
GENERAL INFORMATION:
APPLICANT: Machur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from
Thermotoga Maritima
NUMBER OF SEQUENCE ADDRESSES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/880,729
FILING DATE: 12-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/472,857
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/951,889
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1744
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:


```
Db 329 CATATGATTAATGACAGAAATTTCCCTCGAAGAGTTAACCATGTTTGCATAGAGGCTCTTG 388
QY 254 AAGAGAGACTGGCTGTGCTATTAATATTCATCATGACGAGAGATTAATGATGATCAG 313
Db 389 AGATATATTAAAGATATATCATATAGCAGCATTTTGAAGAACTGTATCAAGAAACCG 448
QY 314 AAGAACCAAGAAAGAAATTTCTTGTCTTTTGAAGAAATTTGCTGATGTTATTAAGACT 373
Db 449 ATAAATACGGCGATGTTTGTGGAATTTGAGACAGATTTGCAAAATTTCTTAAAGATT 508
QY 374 ATCCCGAACTCATTTTGTGAATTTCTGAATGAACCTCAGGAAATCTTACCCGAAA 433
Db 509 ACCCGGAAATCTGTTCTTTTGAATTTACACAGCCTGCTCAGAACTTGCAGGCTGAAA 568
QY 434 AATGATGAAGTACTGCTTGAAGAGCTCTAAAGATTAATGATCAATTAAGCAAAAGCACA 493
Db 569 AATGAAAGCAGCTTTATCCAAAGAGTCTCAAGATTAACAGAGAGCAATCCAAACCGGA 628
QY 494 CTATTAATTAAGCAGCAGCTGAATGGGGGGGTATATCTGCCCTTGAAGAACTGTCTGCC 553
Db 629 TTGTCATTAATGATGCTCCAAAGCTGGGACACATATAGCGAGAGTGAAGTCTTAAATTAG 688
QY 554 CAAAATGGGAAAAAATTTCTATGATTAATCAATTAATCAATCTTGAATTTTACC 613
Db 689 TCAA---CGCAAACGCGATGTTTCTTCCATTAATCAAGAACCTTTTCAAAATTCACAC 745
QY 614 ATCAAGAGCTGAGTGGGTGGAAGGATCTGAGAAATGTTGGGAGAAAGTGGGGATCTC 673
Db 746 ATCAAGGTTGCGGAGTGGGTTAATCCATCCACCTGTTAGGGTTAAGTGAATGGCGAGG 805
QY 674 CAGATGATCAACAAATTTGATAGAAATTCATTTTATAGAAATGGTCAAAAGAA 733
Db 806 AATGGCAATTTAACAATATGAGAAAGTCAATTTCAATACGAGAGTGGGCAAGCAAA 865
QY 734 ACAAAGACCAATTTACATAGTGAGTGTGCTGCTACAAAGAGTGAATTTGATCA 793
Db 866 ATAACTGACCAATCTTCTGGAATTTGGTCTTAATTCAGAAAGCAGATGGATCCAA 925
QY 794 GAATTAATGAGCTCTTGTGCTTGGCGAAATGAGAAAGAGATGAGCTGGGCT 853
Db 926 GGGTTAAGTGCAGCAAGAGTGTGGAAGAAATGGCGAAGATTTGGATTTTCAATACCGCT 965
QY 854 ACTGGGATTTTGTCCGTTTGTGCTTATGATTAATCTGAGAAAGCTGGATTAAG 913
Db 986 ATGGGATTTTGTGCGATTTGCGATTAATGATAGATGCTGCTCAAACTGATCGAAC 1045
QY 914 ATCTTTAGAACTTATATAGGAGAG 940
Db 1046 CATTGCAACAGCTGTGTTGGCAGAG 1072

RESULT 3
US-09-376-045-5
; Sequence 5, Application US/09376045
; Patent No. US20020019998A1
; GENERAL INFORMATION:
; APPLICANT: Aventis CropScience GmbH
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF
; TITLE OF INVENTION: PLANTS WITH CHANGED SUCROSE CONCENTRATION
; FILE REFERENCE: 514413-3772
; CURRENT APPLICATION NUMBER: US/09/376,045
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Solanum tuberosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (118)..(2841)
US-09-376-045-5
```

```
Query Match 4.2%; Score 40; DB 10; Length 2930;
Best Local Similarity 55.9%; Pred. No. 1.9;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 225 AGTGGATGAGTGTATTAACGAGCCCTGAAAAGAGAGCTGGCTGTCTTAATATTTCA 284
Db 1440 AGCTTATGCTCTCCCATGATGATGAGCAAAAATGAGAGACCTGTTATATATAGGCT 1499
QY 285 TCATACGAGAGTATTAATGATGATCCAGAAAGACATAGAAAGATTTCTGCTCTTTG 344
Db 1500 TCTTGAATATGCTCTTCTTATGATGATCCCATGATCAGAGGCAATTCCTATGCTCTTT 1559
QY 345 GAAACAAATTCCTGAT 360
Db 1560 GAAGTTGTTGCTGAT 1575
```

```
RESULT 4
US-09-376-045-3
; Sequence 3, Application US/09376045
; Patent No. US20020019998A1
; GENERAL INFORMATION:
; APPLICANT: Aventis CropScience GmbH
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF
; TITLE OF INVENTION: PLANTS WITH CHANGED SUCROSE CONCENTRATION
; FILE REFERENCE: 514413-3772
; CURRENT APPLICATION NUMBER: US/09/376,045
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3625
; TYPE: DNA
; ORGANISM: Solanum tuberosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (121)..(3282)
US-09-376-045-3
```

Query Match 4.2%; Score 40; DB 10; Length 3625;

Best Local Similarity 55.9%; Pred. No. 2.2;

Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
QY 225 AGTGGATGAGTGTATTAACGAGCCCTGAAAAGAGAGCTGGCTGTCTTAATATTTCA 284
Db 1881 AGCTTATGCTCTCCCATGATGATGAGCAAAAATGAGAGACCTGTTATATATAGGCT 1940
QY 285 TCATACGAGAGTATTAATGATGATCCAGAAAGACATAGAAAGATTTCTGCTCTTTG 344
Db 1941 TCTTGAATATGCTCTTCTTATGATGATCCCATGATCAGAGGCAATTCCTATGCTCTTT 2000
QY 345 GAAACAAATTCCTGAT 360
Db 2001 GAAGTTGTTGCTGAT 2016

RESULT 5
US-09-376-045-1
; Sequence 1, Application US/09376045
; Patent No. US20020019998A1
; GENERAL INFORMATION:
; APPLICANT: Aventis CropScience GmbH
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF
; TITLE OF INVENTION: PLANTS WITH CHANGED SUCROSE CONCENTRATION
; FILE REFERENCE: 514413-3772
; CURRENT APPLICATION NUMBER: US/09/376,045
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3740
; TYPE: DNA
; ORGANISM: Solanum tuberosum
```

[illegible]

Db 1584 AAGATCTTGAGTGAATAATTATGAAGATCTCATGATGTTTATTTTGAAGG 1525
 Oy 838 AGAT 841
 Db 1524 AAT 1521

RESULT 9

US-09-917-800A-1443
 ; Sequence 1443 Application US/09917800A
 ; Patent No. US20020119462A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Mendrick, Donna
 ; APPLICANT: Porter, Mark
 ; APPLICANT: Johnson, Kory
 ; APPLICANT: Castle, Arthur
 ; APPLICANT: Elashoff, Michael
 ; APPLICANT: Gene Logic, Inc.
 ; TITLE OF INVENTION: Molecular Toxicology Modeling
 ; FILE REFERENCE: 44921-5038-US
 ; CURRENT APPLICATION NUMBER: US/09/917,800A
 ; CURRENT FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: US 60/222,040
 ; PRIOR FILING DATE: 2000-07-31
 ; PRIOR APPLICATION NUMBER: US 60/222,880
 ; PRIOR FILING DATE: 2000-11-02
 ; PRIOR APPLICATION NUMBER: US 60/290,029
 ; PRIOR FILING DATE: 2001-05-11
 ; PRIOR APPLICATION NUMBER: US 60/290,645
 ; PRIOR FILING DATE: 2001-05-15
 ; PRIOR APPLICATION NUMBER: US 60/292,336
 ; PRIOR FILING DATE: 2001-05-22
 ; PRIOR APPLICATION NUMBER: US 60/295,798
 ; PRIOR FILING DATE: 2001-06-06
 ; PRIOR APPLICATION NUMBER: US 60/297,457
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: US 60/298,884
 ; PRIOR FILING DATE: 2001-06-19
 ; PRIOR APPLICATION NUMBER: US 60/303,459
 ; PRIOR FILING DATE: 2001-07-09
 ; NUMBER OF SEQ ID NOS: 1740
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 1443
 ; LENGTH: 3378
 ; TYPE: DNA
 ; ORGANISM: Rattus norvegicus
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020119462A1 U01914
 US-09-917-800A-1443

Query Match 3.9%; Score 36.8; DB 10; Length 3378;
 Best Local Similarity 56.7%; Pred. No. 15;
 Matches 68; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Oy 252 GAAAGAGAGCTGGCTGCTTAATAATTCATCTACAGAGATTAAATGATGCC 311
 Db 1542 GAAAGAGAGCTGGCTGCTTAATAATTCATCTACAGAGATTAAATGATGCC 311
 Oy 312 AGAAGACACAGAGAGATTCTTCTGTAACAACAATTCGATGTTATAACA 371
 Db 1602 AGAAGACACAGAGAGATTCTTCTGTAACAACAATTCGATGTTATAACA 371

RESULT 10

US-09-797-464A-1
 ; Sequence 1, Application US/09797464A
 ; Publication No. US20030022807A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wiltling, Reinhard
 ; APPLICANT: Bjornvad, Mads Eskelund
 ; APPLICANT: Kauppinen, Markus Sakari
 ; APPLICANT: Schulten, Martin
 ; TITLE OF INVENTION: Family 5 Xyloglucanases

; FILE REFERENCE: 6073.200-US
 ; CURRENT APPLICATION NUMBER: US/09/797,464A
 ; CURRENT FILING DATE: 2002-02-19
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: Patentln version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 2392
 ; TYPE: DNA
 ; ORGANISM: Paenibacillus pabuli
 US-09-797-464A-1

Query Match 3.8%; Score 36.6; DB 9; Length 2392;
 Best Local Similarity 49.7%; Pred. No. 14;
 Matches 93; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Oy 99 GATRAAAGATGATGTTCTTCGACATTTAAGAAGCCGGTTCTCTCATGTTGCAATCC 158
 Db 950 GGTAACTCCAGAGTTATCAAAAAGTAAAGCGCGCTTCAATCCATTCATTC 1009
 Oy 159 AATAAGATGAGTACGACCGCTTCCCTTATTAATCATGGATCGCTCTT 218
 Db 1010 CGTTTCTTATTTGACACATTTGGAAGCGCTCCCATTTATACATTTATGCGCATGGCT 1069
 Oy 219 CAAAAGATGATGATGATTAACGAGCCCTGAAAAGAGAGCTGCTGTCTATAAA 278
 Db 1070 GAATCGAATTCAGCAAGTCTGACCTATCGGTACATGAAGGCTGTATGATCATCA 1129
 Oy 279 TATTCAT 285
 Db 1130 TATTCAT 1136

RESULT 11

US-09-918-995-15849/c
 ; Sequence 15849, Application US/09918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918,995
 ; CURRENT FILING DATE: 2001-07-30
 ; PRIOR APPLICATION NUMBER: US/09/235,076
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 15849
 ; LENGTH: 455
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)...(455)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-918-995-15849

Query Match 3.8%; Score 36.4; DB 9; Length 455;
 Best Local Similarity 56.8%; Pred. No. 6.2;
 Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Oy 368 AAGACATCCCGAAGCTATTTTGAATTCGATGAACCTCAGCAATCTTCTC 427
 Db 265 AAGTGAACAGGTTATTTGATTTCTTTATGAAGATTAAGTGAAGACATGATGTC 206
 Oy 428 CGGAAAAATGGAATGAACTCTTGAGAGACCTTAAGATTATAGATCAATTGACAA 485
 Db 205 TGGAAGGTGAAGTAATCTGAGAGACCTCAAAAAGCTATTAATTTGTGGAGAA 148

RESULT 12

US-09-822-849A-492/c
 ; Sequence 492, Application US/09822849A

[illegible]

	Query Match	3.8%;	Score 36.4;	DB 10;	Length 2447;
	Best Local Similarity	56.8%;	Pred. No. 16;		
	Matches 67;	Conservative	0;	Mismatches 51;	Indels 0;
				Gaps	
Oy	368	AAGACTATCCCGAAGCTCATTTTGGAAATCTCAATGAACCTCAGCGAAATCTTACTC	427		
Db	495	AAGGAAACAGGTTATGGGATTTTCTTATGAGAAATAAAGTGGAGAAACGATAGTC	436		
Oy	428	CGGAAAAATGGAATGAACTGCTTGACGAGACTCTAAAAGTTATAGATCAATTGACAA	485		
Db	435	TGGAAAGGTGAAGTAATCTGCATAGCGCGTACCAAAAGGATATAAATTTGTGTGAGAA	378		

Search completed: April 26, 2003, 10:58:01
Job time : 129 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 13:49:58 ; Search time 40 Seconds
(without alignments)
1056.011 Million cell updates/sec

Title: US-09-880-729a-2
Perfect score: 1732
Sequence: 1 MGVDPERNKLIGRCINIGN.....LRKTWNKDLLEALLIGDSITE 317

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq.101002:*

- 1: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 23: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1732	100.0	317	20	AAV39338 Carboxymethyl cell
2	1732	100.0	317	20	AAV39338 Carboxymethyl cell
3	1732	100.0	317	21	AAV39338 Carboxymethyl cell
4	1732	100.0	317	22	AAE08543 Thermotoga maritima
5	835.5	48.2	360	19	AAW49870 Thermotoga OC1/4V
6	812.5	46.9	360	18	AAW34566 Thermotoga OC1/4V
7	315	18.2	452	15	AAW49102 Translated sequenc
8	314	18.1	800	15	AAW47496 Translated sequenc
9	312	18.0	477	21	AAW08774 Amino acid sequenc
10	309	17.8	406	11	AAW08199 Neutral cellulase

11	308	17.8	390	22	AAW62468
12	308	17.8	394	22	AAW62471
13	308	17.8	582	22	AAW62467
14	306.5	17.7	574	17	AAW00383
15	298.5	17.2	329	19	AAW35002
16	296.5	17.1	471	19	AAW56742
17	292	16.9	395	22	AAE09784
18	285	16.5	400	22	AAE09785
19	283.5	16.4	348	22	AAE09788
20	281.5	16.3	537	20	AAW8463
21	279	16.1	363	22	AAE09791
22	264.5	15.3	566	12	AAW13228
23	203	11.7	160	22	AAE09789
24	203	11.7	167	22	AAE09786
25	181	10.5	410	18	AAW12378
26	181	10.5	411	18	AAW12379
27	181	10.5	411	18	AAW12381
28	181	10.5	412	18	AAW12380
29	179	10.3	409	9	AAW1843
30	177	10.2	467	17	AAW05731
31	177	10.2	467	17	AAW05731
32	170	9.8	352	20	AAW08472
33	169.5	9.8	369	20	AAW14150
34	168	9.7	400	18	AAW23601
35	168	9.7	400	18	AAW22521
36	168	9.7	400	19	AAW57431
37	168	9.7	462	19	AAW57433
38	165	9.5	472	19	AAW37243
39	163.5	9.4	499	14	AAW42122
40	163	9.4	403	23	AAW81820
41	163	9.4	405	21	AAW43848
42	159	9.2	551	18	AAW18790
43	158.5	9.2	476	19	AAW37241
44	156.5	9.0	484	19	AAW43910
45	155	8.9	386	19	AAW37242

ALIGNMENTS

RESULT 1
AAV39338
ID AAV39338 standard; Protein; 317 AA.
AC AAV39338;
DT 01-DEC-1999 (first entry)
DE Carboxymethyl cellulase (CMCase).
XX
XX
XX Carboxymethyl cellulase: CMCase; thermostable enzyme; hydrolysis:
KW beta 1,4 glycosidic bond; cellulose; plant biomass conversion; fuel;
KW chemical; detergent; textile industry; animal feed; waste treatment;
KW fruit juice extraction; clarification; brewing industry.
XX
XX Thermotoga maritima.
OS
XX
XX US5962258-A.
XX
XX 05-OCT-1999.
XX
XX 23-AUG-1995; 95US-0518615.
XX
XX 23-AUG-1995; 95US-0518615.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Mathur EJ, Lam DE;
XX
XX MPI: 1999-571269/48.
XX
XX N-PSDB: AA221563.
XX
XX Recombinant polynucleotide sequence encoding a carboxymethyl cellulase

PF useful for the conversion of plant biomass -
 XX Claim 1; Fig 1; 14pp: English.

PS This is the amino acid sequence of the Thermotoga maritima carboxymethyl
 CC cellulase (CMCase). CMCase is a thermostable enzyme (optimum temperature
 CC of 60 + degrees C), that catalyses the hydrolysis of the beta 1,4
 CC glycosidic bonds in cellulose. The enzyme degrades cellulose by the
 CC hydrolysis of this bond. The CMCase enzyme has a molecular weight of
 CC approximately 35kD. CMCase is used to catalyse enzymatic degradation of
 CC cellulose and can be used for the conversion of plant biomass into fuels
 CC and chemicals. CMCase can also be used in detergents, the textile
 CC industry, animal feed, waste treatment and for the clarification and
 CC extraction of juices in the fruit juice and brewing industries. The
 CC CMCase gene and 15 bp fragments of it can be used as probes to isolate
 CC nucleic acid molecules encoding CMCase enzymes from genomic libraries or
 CC to screen for related sequences. Cells expressing the CMCase
 CC polynucleotide can be used as immunogens to produce antibodies which can
 CC then be used to isolate the enzyme from cells which express CMCase. The
 CC antibodies can also be used to screen for similar enzymes from other
 CC organisms and samples.

XX Sequence 317 AA:

Query Match 100.0%; Score 1732; DB 20; Length 317;
 Best Local Similarity 100.0%; Pred. No. 3.8e-149; Indels 0; Gaps 0;
 Matches 317; Conservative 0; Mismatches 0;

QY 1 MGVDPERKKIIGRGINIGNALFAPNEGDMGVVINKDEFDIIKEAGFSHVRIPIRMSTHA 60
 DB 1 MGVDPERKKIIGRGINIGNALFAPNEGDMGVVINKDEFDIIKEAGFSHVRIPIRMSTHA 60
 QY 61 YAFPPYKIMDRFRRVDEVINGALKRGLAVAINIHHEELMNDPEHKEKRLALMKQIAD 120
 DB 61 YAFPPYKIMDRFRRVDEVINGALKRGLAVAINIHHEELMNDPEHKEKRLALMKQIAD 120
 QY 121 RYKDYBETLFEFLINPHEGHLTPPEKNELLEALKVIRSIDKKHTIIIGTAEWGGSIALE 180
 DB 121 RYKDYBETLFEFLINPHEGHLTPPEKNELLEALKVIRSIDKKHTIIIGTAEWGGSIALE 180
 QY 181 KLSVPKMEKNSIYTIHYNPFETHGAGAEWEGSEKWLGRKWSGSPDDOKHLEEFNFTIEE 240
 DB 181 KLSVPKMEKNSIYTIHYNPFETHGAGAEWEGSEKWLGRKWSGSPDDOKHLEEFNFTIEE 240
 QY 241 WSKKNKRPYIIGFEGAYRRADLESRIKWTSPVVREREMKRRMSMAWEPFCSGFYVDTLRK 300
 DB 241 WSKKNKRPYIIGFEGAYRRADLESRIKWTSPVVREREMKRRMSMAWEPFCSGFYVDTLRK 300
 QY 301 TWNKDLLEALIGDSIE 317
 DB 301 TWNKDLLEALIGDSIE 317

RESULT 2
 ID AAY23764 standard; Protein: 317 AA.

AC AAY23764;

DT 10-SEP-1999 (first entry)

DE A carboxymethyl cellulase enzyme.

KW Carboxymethyl cellulase enzyme; hydrolysis; cellulose; plant biomass;

KW fuel; chemical; detergent; textile industry; animal feed;

KW waste treatment; fruit juice; brewing industry; clarification;

OS Thermotoga maritima.

PN US9525749-A.

PD 20-JUL-1999.

XX 23-AUG-1995; 9505-0518615.

XX 23-AUG-1995; 9505-0518615.

XX 24-APR-1998; 9805-0066075.

XX (DIVE-) DIVERSA CORP.

XX Lam DE, Mathur EJ;

XX WPI: 1999-429058/36.

XX N-PSDB: AAX85879.

XX New carboxymethyl cellulase enzyme is useful in cellulose digestion

XX Disclosure: Fig 1A-C; 13pp: English.

XX The present sequence represents a carboxymethyl cellulase enzyme from
 CC Thermotoga maritima. Oligonucleotide probes derived from the
 CC polynucleotide sequence can be used to screen a library of genomic
 CC DNA for recovery of the polynucleotide, as PCR primers to synthesise
 CC full length polynucleotides or as diagnostic probes useful for
 CC screening for similar enzymes from other organisms and samples.
 CC The carboxymethyl cellulase enzymes can be used to catalyse hydrolysis
 CC of cellulose for the conversion of plant biomass into fuels and
 CC chemicals, for use in detergents, the textile industry, in animal feed,
 CC waste treatment and in the fruit juice/brewing industry for the
 CC clarification and extraction of juices.

XX Sequence 317 AA:

Query Match 100.0%; Score 1732; DB 20; Length 317;
 Best Local Similarity 100.0%; Pred. No. 3.8e-149; Indels 0; Gaps 0;
 Matches 317; Conservative 0; Mismatches 0;

QY 1 MGVDPERKKIIGRGINIGNALFAPNEGDMGVVINKDEFDIIKEAGFSHVRIPIRMSTHA 60
 DB 1 MGVDPERKKIIGRGINIGNALFAPNEGDMGVVINKDEFDIIKEAGFSHVRIPIRMSTHA 60
 QY 61 YAFPPYKIMDRFRRVDEVINGALKRGLAVAINIHHEELMNDPEHKEKRLALMKQIAD 120
 DB 61 YAFPPYKIMDRFRRVDEVINGALKRGLAVAINIHHEELMNDPEHKEKRLALMKQIAD 120
 QY 121 RYKDYBETLFEFLINPHEGHLTPPEKNELLEALKVIRSIDKKHTIIIGTAEWGGSIALE 180
 DB 121 RYKDYBETLFEFLINPHEGHLTPPEKNELLEALKVIRSIDKKHTIIIGTAEWGGSIALE 180
 QY 181 KLSVPKMEKNSIYTIHYNPFETHGAGAEWEGSEKWLGRKWSGSPDDOKHLEEFNFTIEE 240
 DB 181 KLSVPKMEKNSIYTIHYNPFETHGAGAEWEGSEKWLGRKWSGSPDDOKHLEEFNFTIEE 240
 QY 241 WSKKNKRPYIIGFEGAYRRADLESRIKWTSPVVREREMKRRMSMAWEPFCSGFYVDTLRK 300
 DB 241 WSKKNKRPYIIGFEGAYRRADLESRIKWTSPVVREREMKRRMSMAWEPFCSGFYVDTLRK 300
 QY 301 TWNKDLLEALIGDSIE 317
 DB 301 TWNKDLLEALIGDSIE 317

RESULT 3
 ID AAY56814 standard; Protein: 317 AA.

AC AAY56814;

DT 31-MAR-2000 (first entry)

DE T. maritima thermostable CMCase enzyme.

XX Endoglucanase; carboxymethyl cellulase; cellulose; CMCase; plant biomass;

KW beta-1,4-glycosidic bond; detergent; textile; guar gum; animal feed;

KW waste product; fuel; brewing; thermostable.


```

QY 126 PTEFEELNEPHGULPEKKNELLEELKATRSIDKHTIIOTABMGISALEKLSVP 165
    ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 PENLFEELYNEBPAONLTAKEKNALLPKVLKVIRESNPTRIYIIDAPWMAHSAVRSKL 230
QY 186 KKEKNSIYTHIHYNPEFETHOGAEWVEG---SEKWLGRKKSPPDOCKHLEEFNEFEW 241
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 N-DKIIYFHYEPFKRTHOGAEWVNPIDPVRRKKNMGEEELIQLNISH----FKYSDW 285
QY 242 SKKNRPITYICEFGAYRKADLESRIKMTSEYVREMEKRRMSMAWEFCSGFGYDYLTKRT 301
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 AKQNNVPITFLEEFGAYSKADMSRVKMTESVRKMAEEFGESYAWFEYAGFGIYDRWSON 345
QY 302 WNKDLLEALIG 312
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 WIEPLATAVVG 356

RESULT 7
AAR49102
ID AAR49102 standard; Protein: 452 AA.
XX
AC AAR49102;
XX
DF 21-JUN-1994 (first entry)
XX
DE Translated sequence of domain III of celd cDNA in clone pCNP4.
XX
KW Cellulase; celd; pCNP4; anaerobic rumen.
XX
OS Neocallimastix patriciarum.
XX
PN W09400578-A.
XX
PD 06-JAN-1994.
XX
PE 24-JUN-1993; 93WO-AU00307.
XX
PR 24-JUN-1992; 92AU-0003096.
XX
PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
PI Aylward JH, Gobius KS, Orpin CG, Xue GP.
XX
DR WPI: 1994-026214/03.
XX
N-PSDB; AA055036.
XX
PT Cloning of cellulase clones from anaerobic rumen - by isolating
PT mRNA from culture, converting to cDNA etc. fungi, producing
PT enzymes useful in food processing etc., and DNA for modifying
PT rumen or silage bacteria
XX
PS Claim 10; Page 46-48; 71pp; English.
XX

Clone pCNP4.1 encodes celd. It was derived as follows. N.
patriciarum was used to construct a cDNA library in ZAPII (in E.
coli). Transformants were selected for enzyme activity using
selective media. 11 colonies were positive, and of these 10 had the
same restriction pattern, and the longest of these was designated
celd (pCNP4.1)(AA055036)(AAR47496,R49102). A similar process was used
to isolate the xylanase clone pMX-Tac (AA055037,R47497). An enzyme
composition contg. celd and xylanase is claimed.
XX
SQ Sequence 452 AA;

Query Match 18.28; Score 315; DB 15; Length 452;
Best Local Similarity 27.48; Pred. No. 4.3e-20;
Matches 87; Conservative 56; Mismatches 120; Indels 54; Gaps

QY 10 KILRGINIGINALEA-----PNEGDMG-VVIRDEFDIIKEAGFSHVRPIR 55
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 18 KEMFGNMLGTLDAQCLIEYLYNKDQTASFTCGNENKTTEDMEFKVYMDQFNVRIPPT 77
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 56 WSTHAAYAFPPYKINDREFKRVDEVINGALRGIAVALINHH-----YEELMDPBEHKE 109

```

[illegible]

CC polysaccharide containing a beta-1,3' or beta-1,4' glycosidic linkage.
CC The egla polynucleotide is useful for producing cellulase enzymes that
CC can hydrolyze the glycosidic linkages in polysaccharides such as
CC cellulose. The enzyme is also useful in a number of industrial
CC applications where breaking down biomass is beneficial. Cellulase is also
CC useful as a supplement in animal feed to decrease the production of fecal
CC waste by increasing the digestibility of the feed. Cellulase is also
CC useful to increase the efficiency of alcoholic fermentations by
CC converting undigestible biomass into fermentable sugars. The present
CC sequence represents the catalytic domain of egla polypeptide.

xx Sequence 390 AA;

Query Match 17.8%; Score 308; DB 22; Length 390;
Best Local Similarity 26.3%; Pred. No. 1.5e-19;
Matches 83; Conservative 54; Mismatches 127; Indels 52; Gaps 8;

QY 10 KILRGHIGNALA-----PNEGDMC-VIKREFDIIKEAGFSNRIPIR 55
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 39 KEMNGWMLGNTMDAQCIETLYNTERKDQTASFCWCNPKTTEDMFKVLIDNQFNFRIPIT 98

QY 56 WSTHAAYAPPPKIMDRFPRVDEVINGALRKGLAVAINIH-----YEELMNDPEEHE 109
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 99 WSGHGGEAPDRKIDEKWLAKRYHEVVDDYYKKGAAYIILNHETNNHASESLDLRAKELLE 158

QY 110 RFLALMKOIAIRXDYPETLEFEELINEPHGNLTPEKW-----NEELEALKYIR 158
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 159 K---TMSQIAEEFKYDEHLIFEGLINEPRKNDTPVEWTGDQEGWDVANMMNAVPLKTVR 215

QY 159 SI----DKKHIIITICTAWMGDISALEKLSVKWKENKSIVTHIHYNPFETFOGAENWGCS 214
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 216 SAGGNPNRRHLMPYAACNENSFNFIPEDDDKVASVAATPYFALNNGGAVDK 275

QY 215 EKMVLGRK--WGSPDDOKHLIEEFNFIEEWSKKNRPTIYGEPGARADLESRIKMTSF 271
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 276 FDAGKRRLLENINIMKKRFVDQ-----GIRPILDEYGMANDNEEDRATMAEF 324

QY 272 VRENERKRRMSWAYWE 287
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 325 YMEKYTAGVPQIWMD 340

RESULT 12
AAB62471
ID AAB62471 standard; Protein: 394 AA.

XX AC AAB62471;
XX DT 09-JUL-2001 (first entry)
XX DE P. rhizinflata egla protein partial sequence.
XX KW Cellulase; egla: fungus; polysaccharide; glycosidic linkage; enzyme;
XX KM animal feed; alcohol; fermentation.
XX XX Piromyces rhizinflata.
XX OS US6222028-B1.
XX PN 24-APR-2001.
XX PD 15-OCT-1999; 99US-0419459.
XX PF 15-OCT-1999; 99US-0419459.
XX PR 15-OCT-1999; 99US-0419459.
XX PA (STINT-) ACAD SINICA.
XX PI Liu J., Cheng K., Tsai C., Chang C;
XX DR MPI: 2001-307598/32.
XX N-PSDB: AAF83296.
XX New isolated nucleic acid encoding a cellulase enzyme of fungus

PT Piromyces rhizinflata, for producing cellulase that hydrolyzes
 PT polysaccharides containing beta-1,3 or beta-1,4 glycosidic linkages
 XX
 PS Disclosure; Columns 7-8; 16pp; English.
 XX
 CC The invention relates to a cellulase enzyme, egIA, isolated from the
 CC fungus Piromyces rhizinflata. The egIA polypeptide hydrolyzes a
 CC polysaccharide containing a beta-1,3' or beta-1,4' glycosidic linkage.
 CC The egIA polynucleotide is useful for producing cellulase enzymes that
 CC can hydrolyze the glycosidic linkages in polysaccharides such as
 CC cellulose. The enzyme is also useful in a number of industrial
 CC applications where breaking down biomass is beneficial. Cellulase is also
 CC useful as a supplement in animal feed to decrease the production of fecal
 CC waste by increasing the digestibility of the feed. Cellulase is also
 CC useful to increase the efficiency of alcoholic fermentations by
 CC converting indigestible biomass into fermentable sugars. The present
 CC sequence represents an amino acid sequence of an egIA catalytic domain
 CC downstream of the GSTin an GST-egIA fusion protein.
 XX
 SQ Sequence 394 AA:
 Query Match 17.8%; Score 308; DB 22; Length 394;
 Best Local Similarity 26.3%; Pred. No. 1.6e-19;
 Matches 83; Conservative 54; Mismatches 127; Indels 52; Gaps 8;
 OY 10 KILGRGINGNALAE-----PNEGDMG-VVTKDEFDIIKAGFSHVRIPIR 55
 DB 43 KENFGMNLGNTMDACIEYLVNKKDQTASFTGCMGPKTTEDEMFVLIDNQFVFRIPPT 102
 OY 56 WSHVAVAFPPYKIMDFEKKRVDEVINGALKRGLAVAINHH-----YELMDPEEKE 109
 DB 103 WSGHFGADPYKIDKWLKRVHEVDYPRKNGAFVILNLHETWNAFSETLDIAKEILE 162
 OY 110 RFLAKKOIADRYKDYPTLFEFLNEPHGNLTPEKW-----NELLEALKVIR 158
 DB 163 K--IMSQIAEEFKDYDEHLIFEGLINEPRKNDTPVEWTGDOGDMVANNMNAVFLKTVR 219
 OY 159 SI-----DKKHTITIGTAEMGISALEKLSVPKMEKNSIVTHIYNPPEFHOGAEWEGS 214
 DB 220 SAGGNPKRRLMIPYAACNENSPNNFTPEDDOKYIASVHAYAPNPLNNGGAVDK 279
 OY 215 EKMLGKR--WGSPPDQKHLEEFNFIEEMSKKRRPIYIGFGAVRKADLESRIKTSF 271
 DB 280 FDAAGKRDLDMNINLMKKRRVDQ-----GIPMLIGEGAMNRDNEEDRATWAEF 328
 OY 272 VVREMEKRRMSWAYVE 287
 DB 329 YMEKVTAMGVPQIWM 344
 RESULT 13
 AAB62467
 ID AAB62467 standard; Protein: 582 AA.
 XX
 AC AAB62467;
 XX
 DT 09-JUL-2001 (first entry)
 XX
 DE P. rhizinflata egIA protein partial sequence.
 XX
 KW Cellulase; egIA; fungus; polysaccharide; glycosidic linkage; enzyme;
 KW animal feed; alcohol; fermentation.
 XX
 OS Piromyces rhizinflata.
 XX
 FH Key
 FT Region 1..109 Location/Qualifiers
 FT /note= "repeat region (AAB62470)"
 FT Domain 110..499
 FT /note= "catalytic domain (AAB62468)"
 FT Region 391..499
 FT /note= "repeat region (AAB62471)"
 XX

PN US6222028-B1.
 XX
 PD 24-APR-2001.
 XX
 PF 15-OCT-1999; 99US-0419459.
 XX
 PR 15-OCT-1999; 99US-0419459.
 XX
 PA (SINI-) ACAD SINICA.
 PI Liu J, Cheng K, Tsai C, Chang C;
 XX
 XX WPI: 2001-307598/32.
 DR N-PSDB: AAF83292.
 XX
 PT New isolated nucleic acid encoding a cellulase enzyme of fungus
 PT Piromyces rhizinflata, for producing cellulase that hydrolyzes
 PT polysaccharides containing beta-1,3 or beta-1,4 glycosidic linkages
 XX
 PS Disclosure; Columns 5-6; 16pp; English.
 XX
 CC The invention relates to a cellulase enzyme, egIA, isolated from the
 CC fungus Piromyces rhizinflata. The egIA polypeptide hydrolyzes a
 CC polysaccharide containing a beta-1,3' or beta-1,4' glycosidic linkage.
 CC The egIA polynucleotide is useful for producing cellulase enzymes that
 CC can hydrolyze the glycosidic linkages in polysaccharides such as
 CC cellulose. The enzyme is also useful in a number of industrial
 CC applications where breaking down biomass is beneficial. Cellulase is also
 CC useful as a supplement in animal feed to decrease the production of fecal
 CC waste by increasing the digestibility of the feed. Cellulase is also
 CC useful to increase the efficiency of alcoholic fermentations by
 CC converting indigestible biomass into fermentable sugars. The present
 CC sequence represents a partial egIA polypeptide.
 XX
 SQ Sequence 582 AA:
 Query Match 17.8%; Score 308; DB 22; Length 582;
 Best Local Similarity 26.3%; Pred. No. 2.7e-19;
 Matches 83; Conservative 54; Mismatches 127; Indels 52; Gaps 8;
 OY 10 KILGRGINGNALAE-----PNEGDMG-VVTKDEFDIIKAGFSHVRIPIR 55
 DB 148 KENFGMNLGNTMDACIEYLVNKKDQTASFTGCMGPKTTEDEMFVLIDNQFVFRIPPT 207
 OY 56 WSHVAVAFPPYKIMDFEKKRVDEVINGALKRGLAVAINHH-----YELMDPEEKE 109
 DB 208 WSGHFGADPYKIDKWLKRVHEVDYPRKNGAFVILNLHETWNAFSETLDIAKEILE 267
 OY 110 RFLAKKOIADRYKDYPTLFEFLNEPHGNLTPEKW-----NELLEALKVIR 158
 DB 268 K--IMSQIAEEFKDYDEHLIFEGLINEPRKNDTPVEWTGDOGDMVANNMNAVFLKTVR 324
 OY 159 SI-----DKKHTITIGTAEMGISALEKLSVPKMEKNSIVTHIYNPPEFHOGAEWEGS 214
 DB 325 SAGGNPKRRLMIPYAACNENSPNNFTPEDDOKYIASVHAYAPNPLNNGGAVDK 384
 OY 215 EKMLGKR--WGSPPDQKHLEEFNFIEEMSKKRRPIYIGFGAVRKADLESRIKTSF 271
 DB 385 FDAAGKRDLDMNINLMKKRRVDQ-----GIPMLIGEGAMNRDNEEDRATWAEF 433
 OY 272 VVREMEKRRMSWAYVE 287
 DB 434 YMEKVTAMGVPQIWM 449
 RESULT 14
 AAM00383
 ID AAM00383 standard; Protein: 574 AA.
 XX
 AC AAM00383;
 XX
 DT 31-JAN-1997 (first entry)
 XX

```
DE   Bacillus cellulase BCE 113.  
xx  
KM   Cellulase; BCE 113; detergent; surfactant; laundry;  
KM   tensile strength; antipilling.  
xx  
OS   Bacillus sp. strain CBS 669..93.  
xx  
FH   Key Location/Qualifiers  
FT   Peptide      1..26  
ET       /label= Sig_peptide  
FT   Protein      27..574  
ET       /label= Mat_protein  
FT       /note= "the 63 kDa mature protein is used in  
FT       detergent comps."#  
PM   MO9634092-A2.  
xx  
PD   31-OCT-1996.  
xx  
PE   26-APR-1996;    96WO-EPO1755.  
xx  
PR   12-MAR-1996;     96US-0614115.  
PR   28-APR-1995;     95EP-0201115.  
xx  
PA   (GENV ) GENENCOR INT INC.  
xx  
PI   Kottwitz B, Lenting HBM, Maurer K, Van Beckhoven RFC;  
PI   Van Solingen P, Weiss A,  
DR   WPI: 1996-49762/49.  
N-PSDB: AAT41849.  
xx  
PT   Cellulase with low ratio of tensile strength loss to antipilling  
PT properties - used in detergent composition which provides  
PT anti-greying, softening, anti-wrinkling and colour protection to  
PT fabrics  
PS  
PS Claim 10; Fig 4; 33pp; English.  
xx  
CC   A novel cellulase (AAMW0383), designated BCE 113, of alkali-tolerant  
CC Bacillus sp. strain CBS 669..93 has a tensile strength loss to  
CC antipilling ratio below 1. It can be isolated from CBS 669..93  
CC fermentation broth or expressed at high levels in transformed host  
CC cells utilising an isolated gene sequence (AAT41849). Cellulase BCE  
CC 113, and similarly isolated cellulase BCE 103 (see also AAMW0382),  
CC show good activity at alkaline pH, and can be used in laundry  
CC detergent comps. to provide anti-greying, softening, anti-  
CC wrinkling and colour protection to fabrics.
```

Db	269	-DSNRLEIEFDVRHHFVANGIPVVLGSEGLLGFQKHTGVIQGGKAKPEFYLIIHNLN	326
Oy	278	KRRMSVAIWEFCSGFG-----VDTLKRTV	302
Db	327	ERDITHMLMDNGQHFRHRTYEWYDEELFDMLRASW	361
RESULT 15			
	AAM35002		
	AAM35002	standard; Protein, 329 AA.	
XX			
XX	AAM35002;		
XX			
XX	21-MAY-1998	(first entry)	
XX			
DE		Thermotoga maritima endoglucanase.	
XX			
KW		Endoglucanase; cellulase; carboxymethylcellulose; cellulose;	
RW		biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;	
KW		thermostable enzyme; thermophilic; glycosidase.	
XX			
OS		Thermotoga maritima (Clone 6GA2).	
XX			
PN		WO9744361-A1.	
PD			
XX		27-NOV-1997.	
XX			
XX		22-MAY-1997;	97WO-US08793.
XX			
XX		22-MAY-1996;	96US-0651572.
XX			
PA		(RECO-) RECOMBINANT BIOCATALYSIS INC.	
XX			
PI		Lam DE, Mathur EJ;	
XX			
DR		WPI: 1998-018435/02.	
XX			
XX		N-PSDB; AAT94210.	
PT			
PT		Endoglucanase(s), preferably form archaeal bacterium, AEPit 1a -	
XX		useful to degrade carboxymethylcellulose and hydrolyse of	
XX		beta-1,4-glycosidic bonds in cellulose	
PS		Claim 1; Fy9 1R; 164pp; English.	
XX			
CC		This protein comprises an endoglucanase of Thermotoga maritima	
CC		(Clone 6GA2) that is capable of degrading carboxymethylcellulose	
CC		and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It	
CC		has homology to an endoglucanase of archaeobacterium AEPit1a (see	
CC		AAM3485).	
CC		It can be produced from native cells or from recombinant	
CC		host cells, especially prokaryotic host cells transformed with a	
CC		plasmid or virus-derived vector including the endoglucanase DNA	
CC		(see AAT94210).	
CC		24 Endoglucanases (see AAM34986-W35008) are claimed.	
CC		They can be used to degrade cellulose for the conversion of plant	
CC		biomass into fuels and chemicals, for use in detergents, textiles,	
CC		animal feed, waste treatment, and in the fruit juice and brewing	
CC		industries for the clarification and extraction of juices.	
XX			
SQ	Sequence	329 AA;	
Oy	Query Match	17.2%; score 298.5; DB 19; Length 329;	
Db	Best Local Similarity	27.9%; Pred. No. 8.8e-19;	
Oy	Matches	94; Conservative 60; Mismatches 124; Indels 59; Gaps 16;	
Oy	14	GCINIGNALPNEGDWGVIKDEFFDIKAGSGSHVIP--IRKSTHAAYAPPYKIMD 70	
Db	9	RGFNLELFASIKSTGNF---KEEDFLMAOWDFNFVIRIPMCHILMSDRG--NPFIRE 61	
Oy	71	REFKRVDEVINGALKRGVALAINT-----HYEELM--DPEHKKEFLAMKQIA 119	
Db	62	DFFEKIDRIVTWGKEVIGHICISLHRAPGVYKNKEVEKTNLMDETAQEAFIHHWSFTA 121	
Oy	120	DRYVDYDEPT-LFFELINP-----HGNTLPKXWNLLEALVIRSIDKKHHIIIGTAWMG 174	

Db 122 RRYGISSTHLSFNLINEPPFPDQIMSVEDHNSLKRRTITEIRKIDPERLIIDOLGYG 181
 QY 175 GISALEKLSVPKWEKNSIVTIIHYNPFEEFHOGAEWEGSE---KW---LGRKMGSP 225
 Db 182 NI-PVDDLT---ENTVQSCRGYLPFSVTHYKAEMVDSKDPPEWEPNGMHFGYEW--- 233
 QY 226 DDOKHIEEFNFIEEMSKKNR--PIYIOFGAYRRADLESRIKWTSEVVREREMEKRRMSW 283
 Db 234 -NREKLEHY---LWIKLRQKGIIEVCGEMGAYNKTPHDVVLKWLDELLEIFKTLNIGF 289
 QY 284 AYWEFCGFGVYDTLRK-----TW-----NKDLLEAL 310
 Db 290 ALMNFPGPGILDSEKDYEEEMYGHKLDKMLELL 326

Search completed: April 22, 2003, 13:57:57
 Job time : 43 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 13:56:24 ; Search time 45 Seconds
(without alignments)
677.214 Million cell updates/sec

Title: US-09-880-729A-2

Sequence: 1 MGVDPEFRNKILGRGINICN.....LRTWKDLLEALIGDSTIE 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pfr1:*
2: pfr2:*
3: pfr3:*
4: pfr4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1728	99.8	317	2	B72216
2	598	34.5	584	2	J01229
3	598	34.5	900	2	JH0157
4	397	22.9	814	1	CZCLEM
5	377	21.8	370	2	G97001
6	366	21.1	377	2	F97001
7	362	20.9	441	2	A44815
8	356.5	20.6	364	2	S12017
9	346.5	20.0	515	2	S20493
10	341.5	19.7	658	2	A33598
11	330.5	19.1	1012	2	B97326
12	327	18.9	343	2	JE0409
13	324	18.7	343	2	JT0268
14	316.5	18.3	611	2	JC7177
15	315	18.2	1232	2	T31426
16	311.5	18.0	409	2	S12018
17	311.5	18.0	473	2	S40507
18	311	18.0	517	2	I40798
19	309	17.8	406	2	A43722
20	305	17.6	482	2	JE0302
21	298.5	17.2	329	2	C72216
22	297.5	17.2	574	2	C83725
23	287.5	16.6	500	2	S22458
24	285.5	16.5	475	1	CZCLCA
25	269.5	15.6	455	2	S16559
26	265.5	15.3	566	2	JH0218
27	233.5	13.5	547	2	JG0356
28	226.5	13.1	363	2	I40234
29	215.5	12.4	584	2	S27500

30	206.5	11.9	397	2	A35136	cellulase (EC 3.2.
31	182	10.5	482	2	E97012	probable non-proce
32	181	10.5	409	2	B25156	cellulase (EC 3.2.
33	180	10.4	488	2	A25156	cellulase (EC 3.2.
34	173	10.0	338	2	JC4115	endo-1,4-beta gluc
35	172.5	10.0	722	2	H96986	mannan endo-1,4-be
36	169.5	9.8	369	2	T04323	cellulase (EC 3.2.
37	165.5	9.6	915	2	A43802	cellulase (EC 3.2.
38	165.5	9.6	1039	2	S02711	cellulase (EC 3.2.
39	164.5	9.5	429	2	S29044	endo-1,4-beta gluc
40	163.5	9.4	499	2	A27198	cellulase (EC 3.2.
41	163	9.4	403	2	A84592	(1-4)-beta-mannan
42	162.5	9.4	448	2	A27631	cellulase (EC 3.2.
43	160	9.2	410	1	S68153	cellulase (EC 3.2.
44	157.5	9.1	508	2	A26874	cellulase (EC 3.2.
45	156.5	9.0	486	2	I40548	bifunctional cellu

ALIGNMENTS

RESULT 1	B72216	Endoglucanase - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima	C:Date: 11-Jun-1999	#sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: B72216	R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic	Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.	Nature 399, 323-329, 1999	A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316; PMID:10360571	A:Accession: B72216	A:Status: Preliminary
A:Molecule type: DNA	A:Residues: 1-317 <ARN>	A:Cross-references: GB:AE001813; GB:AE000512; NID:g4982321; PIDN:AAD36816.1; PID:g498
A:Experimental source: strain MSB8	C:Genetics: TML751	
Query Match	99.8%; Score:1728; DB 2; Length 317;	
Best Local Similarity	99.7%; Pred. No. 4.5e-117;	
Matches 316; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
OY 1	MGVDPEFRNKILGRGINIGNALPNEGNGVYIKDEFDIIKAGFSHVRIPIRMSTHA 60	
Db 1	MGVDPEFRNKILGRGINIGNALPNEGNGVYIKDEFDIIKAGFSHVRIPIRMSTHA 60	
OY 61	YAPPPKIMDFRFRVDEVINGALRGIAVINIHHEELMNDPEEKERFLAMKQIAD 120	
Db 61	YAPPPKIMDFRFRVDEVINGALRGIAVINIHHEELMNDPEEKERFLAMKQIAD 120	
OY 121	RYVDPEITLFEILNEBHGMLTPKWNLELEALKVRSIDKHTIITGAEMGISALE 180	
Db 121	RYVDPEITLFEILNEBHGMLTPKWNLELEALKVRSIDKHTIITGAEMGISALE 180	
OY 181	KLSVPKWEKNSIYTIHYNNFEETHOGAEVVEGSEKVLGRKMSPPDOKHLEIFENITEE 240	
Db 181	KLSVPKWEKNSIYTIHYNNFEETHOGAEVVEGSEKVLGRKMSPPDOKHLEIFENITEE 240	
OY 241	WSKKNRPITYIGFGAVRKADLESRIKWTSEVVREREMKRRMSNAVMEFGSGFYDTLRK 300	
Db 241	WSKKNRPITYIGFGAVRKADLESRIKWTSEVVREREMKRRMSNAVMEFGSGFYDTLRK 300	
OY 301	TWNKDLLEALIGDSTIE 317	
Db 301	TWNKDLLEALIGDSTIE 317	
RESULT 2	J01229	

Query Match	34.5%;	Score 598;	DB 2;	Length 584;
Best Local Similarity	37.48;	Pred. No. 2e-35;		
Matches 116; Conservative	58;	Mismatches 114;	Indels 22;	Gaps 6;

RESULT 3
JH0157
cellulase (EC 3.2.1.4) H precursor - Clostridium thermocellum
N:alternate names: endo-1,4-beta-glucanase; endoglucanase H
C:Species: Clostridium thermocellum
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 15-Oct-1999
C:Accession: JH0157
R:Yagoue, E.; Beguin, P.; Aubert, J.P.

Gene 89, 61-67, 1990
A:Title: Nucleotide sequence and deletion analysis of the cellulase-encoding gene celH
A:Reference number: JH0157; M0ID:90323606; PMID:2197182
A:Accession: JH0157
A:Molecule type: DNA
A:Residues: 1-900 <TAG>
A:Cross-references: GB:M31903; NID:q144773; PIDN:AAA23225.1; PID:q144774
A:Note: the authors translated the codon CAG for residue 863 as His
A:Comment: Cellulase H is involved in the hydrolysis of cellulose and arranged in cellulose

Query Match	34.58;	Score 598;	DB 2;	Length 900;
Best Local Similarity	36.68;	Pred. No. 3.5e-35;		
Matches 115; Conservative	60;	Mismatches 121;	Indels 18;	Gaps 4.

RESULT 4
CZOLEM
cellulase (EC 3.2.1.4) E precursor - Clostridium thermocellum
N:Alternate names: endo-1,4-beta-glucanase E; endoglucanase E
C:Species: Clostridium thermocellum
C:date: 31-Mar-1992 #sequence-revision 31-Mar-1992 #text-change 18-Jun-1999
C:Accession: J070347
R:Hall, J.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.

A:Title: Conserved reiterated domains in *Clostridium thermocellum* endoglucanases are
Gene 09, 23-56, 1900
A:Reference number: JT0347; MUID:89137992; PMID:3066698

A;Residues: 1814 <M>A;
A;Cross-references: GB:M22759; NID:q144768; PIDN:AAA23224.1; PID:q144770
A;Note: part of this sequence, including the amino end of the mature protein, was con
C;Comment: This secretory enzyme is part of a highly active and thermostable cellulase
C;Genetics:
A;Gene: cele
A;Start codon: GTG
C;Function:
A;Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in beta-D-gluc
A;Pathway: cellulose degradation
C;Superfamily: cellulase CCA; Clostridium cellulase repeat homology
C;Keywords: duplication; extracellular protein; glycosidase; hydrolase; polysaccharid
E;1.34/Domains: signal sequence #status predicted <M>
F;35-814/Product: cellulase E #status predicted <M>
F;415-438/Domains: Clostridium cellulase repeat homology <CCR1>
F;451-474/Domains: Clostridium cellulase repeat homology <CCR2>

[illegible]

Db 282 QNATYNNFVKNGQGVVIGEGSINKNESSASLAKFYVDARKKGGITTYWMD - NGKS 339

QY 292 -----FGYDPLRKTW 302

Db 340 AVGDNDYGGILDRNNLTW 356

RESULT 6

P97001

endoglucanase family 5 [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C:Accession: P97001

R:Nolling, J.; Brecon, G.; Ometchenko, M.V.; Matkarova, K.S.; Zeng, Q.; Gibson, R.; L

.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: P97001

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-377 <KUP>

A:Cross-references: GB:AE001437; PIDN:AAK78801.1; PID:915023716; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC624

C:Genetics:

A:Gene: CAC0825

C:Superfamily: Ruminococcus albus cellulase

Query Match 21.1%; Score 366; DB 2; Length 377;

Best Local Similarity 30.0%; Pred. No. 5,76-19;

Matches 100; Conservative 59; Mismatches 124; Indels 50; Gaps 11;

QY 15 GINGNALFAP-NEGDMG-----VTKDEFEDIKAGFSHWRIPIRWSTHAYAPPY 66

Db 55 GIDLNSLDSPENETDMGNPKTKSMITK-----IKOAGFNVTIRIPITWNSHIGPAPDY 108

QY 67 KIMDFEPRVYDINGALRGGLAVAINIHHEELMNDPEEKERLTA---LMQIADRY 122

Db 109 TIDNAMLDRVQGVVYVADNMHYVLIINHEDSWLIPYVAKDAFALTLRLMQIISNRF 168

QY 123 KDYPELTFFELINPEHGNLTPEKWNELLEALVKVRSI-----DKKHFTII 168

Db 169 KMYNSLSLFIETMNEPRLIGTPEWYGGTSEARADVNSINLAAYNAIRDTGAKNSRFTMI 228

QY 169 GT-AEWGISALEKLSVPRKEKNSIVTITHYNNPEEFTNOGAEWEGSEKWLGRKWSFDD 227

Db 229 PTYASSDSKDISINGLKIPNNDKILISLHAYIPNF-----AMNVNGTS-----SWGSAKD 279

QY 228 OKHLEEFNFIEEMSKKKRPYIGEFAGYKKADESIAKNTSPFYRMEKRRMSMAWE 287

Db 280 KKEIDSTLNATYIKKKGKGGYGVVIGEFASVYKKNMNOBASLAKYVSAKRRKSIAPIWMD 339

QY 288 ---FCSG---EGVYDPLRKTWN-KDLLEALIG 312

Db 340 NCVYSSGAKDSTGIFPNRNTLRMECPKIVHALTG 372

RESULT 7

A44815

cellulase (EC 3.2.1.4) EngB - Clostridium cellulovorans

N:Alternate names: endo-1,4-beta-glucanase EngB

C:Species: Clostridium cellulovorans

C:Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jun-2000

C:Accession: A44815

R:Foong, F.; Hamamoto, T.; Shoseyov, O.; Dol, R.H.

J. Gen. Microbiol. 137, 1729-1736, 1991

A:Title: Nucleotide sequence and characteristics of endoglucanase gene engB from *Clostridium cellulovorans*

A:Reference number: A44815; MUID:92065240; PMID:1955860

A:Accession: A44815

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-441 <FOO>

A>Note: sequence extracted from NCBI backbone (NCBIN:68566, NCBIPI:68570)

	Matches	92;	Conservative	57;	Mismatches	145;	Indels	55;	Gaps	12;
OY	10	KILGCGTGCAGALAPNEDGMGVIKDEFDDIIKAGFSVHRIP	--RWSTHAYAF---	63						
		: : : :	: : : :							
Db	308	KYLKMGVFNTMWLEADCKFKSFELGESDVKILADNGEKSLRLPDLIDLYATNRDAFIAG		367						
OY	64	--PPKYI-MDRFFRKVDVINGALKRGLAVAINIHHEEIMNDPREHKERFLATL----	WK	116						
		: : : :	: : : :							
Db	368	TDTLKKFDDDTLEVLDSFEVEMWTAKYNMSFVIDHYEUNSYTTSAKDPNYIKMAETWK		427						
OY	117	QIADRKYDP-ETTFELLINEP---HGNIITPEKNWELLEALKYISDIKHTIIIGTAIE		172						
		: : : : :	: : : : :							
Db	428	HVAAMAYASSPRDLDFELLNEPDMSDKYTAAITWTTAQAQMIDLAIIRTDYTKHTILLEGDAQ		487						
OY	173	WGGISALEKLSPVKEKNSIVTIHYHNPFEEETHOGAEWE-----GSEKW----	217							
		: : : : :	: : : : :							
Db	488	WYSITLLAK-RTPFDMDNIIVIHTEPEFAFHGGCSTM DYATIHDIPPYDPAKMSFVS		546						
OY	218	--LG-----KKMGSPPDQKHILIEFNFIIEEMSKKNKPITYIGERGAVR-KAD		261						
		: : : : :	: : : : :							
Db	547	GDFGNKSTKSVMKTINIINYRTGSKREALILEOILAKKWAATNNVPVIINEFGALNLRST		606						
OY	262	LESRIKMTSFWREMEKRMRWSWAYEFSCSGFYVTLKTKTWKDLLPAL		310						
		: : : : :	: : : : :							
Db	607	AESRLNLTIAKREICDTLOITPTHWGTGYCNFSYLE-----NGKLIBGI		649						

RESULT 11
B97326
endoglucanase family 5, S-layer homology, cell-adhesion and dockerin domains [imported]
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: B97326
R:NoUling, J.; Berton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: B97326
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1012 <KUR>
A:Cross-references: GB:AF001437; PIRD:AAK81397.1; PIR:G15026560; GSFD:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
C:Gene: CAC3469

```

Query Match          19.1% Score 330.5; DB 2: Length 1012;
Best Local Similarity 28.3%; Pred. No. 7.2e-16;
Matches   98; Conservative    66; Mismatches 131; Indels    51; Gaps    16;

QY      2 GVDPFERNNKILRGINIGNALE-APENEGDMG-VVIKDEFPIIKEAGFSHVRIPIRMSTH 59
       : : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     383 GLSSMDLTCKDKMKVGDNLDSAGSETGWGNPATTRKMIDMQKAGFNVTIRPYRMDH 442
QY      60 AYAPFPYYIMRFPRVRDVEINGALKRGLAVAINHHHEELMNDPEEKERL-----ALW 115
       : : : : : | : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     443 -YTDSNTVIDPAVMSRVTVVYNVALANGMYAIINIHNKHIOGEMNAEKKSDVINEGSAIW 501
QY     116 KQIADRYKDYDEPTELFEEELLNPHGNLTPEKW-----NELLEAKLYRSI---DK 162
       : : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     502 KQVGTHFPRKDYGDOQLFEETINEPR--TGDDMTGNSSYNYVANEYNAKLSVIRAAGNG 558
QY     163 KHTIIIGT----AEMGGISALEKLSVPFKMEKNSTYTHIYNPFETHOGAEVESSEKWL 218
       : : : : : | : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     559 KRLVMPPYCASADAKAAA--MVYPN-DPVVAVSIAHYIPYNA-MNTDTTKSYSTEF 613
QY     219 GRKMSPDDOKHLIE-PNFTIEWSKKNRPIYIEFGAYRADLESRIKMTSFVSRREME 277
       : : : : : | : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     614 G-----DADKACIDKTRLLDKTFIKKGIPVVGGEFAATGKNNNDQNVNFKAKEYLOYVAS 667
QY     278 KRMRSMWAWE-----FCSGEGVYIDLRTKTWN---KULLEALLIGG 313
       : : : : : | : : : : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Db 668 SYGIPCCWMDNNAFSPNSTDSMGIFN--RKTLLQFVYPEIVQAMLDG 711

RESULT 12
J0409
cellulase (EC 3.2.1.4) - Clostridium sp. (strain FI)
N:Alternate names: endo-1,4-beta-glucanase C307; endoglucanase C307
C:Species: Clostridium sp.
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 17-Mar-1995

C:\ncbi\blast\bin\blastn.exe
 R:\Sakka, K.; Shimanuki, T.; Shimada, K.
 Agric. Biol. Chem. 55, 347-350, 1991
 A:Title: Nucleotide sequence of celC307 encoding endoglucanase C307 of Clostridium sp.
 A:Reference number: J0409, MUID:91299276, PMID:1368630

A:Accession: D20757
A:Molecule type: DNA
A:Residues: 1-343 <SAK>
A>Note: part of this sequence was confirmed by protein sequencing
C:Genetics:
A:Gene: celC307

A: description hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
 A: pathway: cellulose degradation
 C: keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match	18.9%; Score 327; DB 2; Length 343;
-------------	-------------------------------------

Matches 97; Conservative 61; Mismatches 125; Indels 62; Gaps 13;

```

QY 15 GINIG---NALEAPNEGDMGVINDEEFEDIKEAGESHVRIPRMTSTHAYAFPPKIMDR 71
    |||:| : : : : ||| |||:|
Db 7 GINLGWISQVQVFSKEHFDFTIEKEDIETIAEAGFPHVRLP-----FDYPIESDN 59

```

```

Qy 72 F-----FRVDEVINGALRGLAVAINIH-----YEELMNDPEEHKERFLAL 114
      : : | : ::| | : | : | :
Db 60 VGEYKEDGLSYIDRCLEWCCKKNLGLVDMHNAFGYRFODFKTSTFEEDPNQK-RFVDI 118

```

Qy 115 WKQIADRYKDYPETLEFFELINPBGNLTP--KNNEELLEALKVRSIDKHTIIIGAE 172

Db 119 WRFIAKRYINERHIAFELLNEY--VEPDSTRNNKLMIECVKAIREIDSTPRMLYIGNN 175

```
Qy 173 WGISALEKISVPKWEKNSIV-TIHYNPEFTHQGAENVEG-----SEKKLGRKWCSP 229
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 176 YNSPDELKNTLA--DIDDDIYVNFHFYNPEFFTHQKAHMSAMAYNRTYVKKPGQYECIE 233
```

```
Qy 226 DDQH-----LIEFNFI-----EWSKKNRPITYGEGAYRAADESRI 266
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 234 EFVKNPNPKYSMMELNNLKLNLKRLKDKPAIEREREKKCKLYGEEFGVIAADESRI 293
```

Qy	267	kwtsfvvrenekrrmswaywefcs-gfcsvydlrktwnkdllleal	310
		: : : : : :	
Db	294	kwHEDYISLLEEDIGAVWNKKMDfELYNEDKKPVSQELVNIL	338

RESULT 13
JT0268

N:Alternate names: endo-1,4-beta-glucanase
C:Species: Clostridium thermocellum
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 15-Oct-1999

R:Schwarz, W.H.; Schimming, S.; Ruecknagel, K.P.; Burschwaiger, S.; Kreil, G.; Staudt Gene 63, 23-30, 1988

A:Accession: J10268
A:Molecule type: DNA
A:Residues: 1-343 <SCH>

A; Note: part of this sequence, including the amino end of the mature protein, was contributed by Dr. J. Wang, Q.; Tull, D.; Melnick, A.; Gillies, N.R.; Warren, A.J.; Aebbersold, R.; Withers, J. Biol. Chem. 268, 14096-14102, 1993

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 13:50:19 ; Search time 14 Seconds

(Without alignments)
939.143 Million cell updates/sec

Title: US-09-880-729A-2

Perfect score: 1732
Sequence: 1 MGVDPEERKNIKRGINICN.....LRKTNKDLLEALIGDSIE 317

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	598	34.5	584	1 GUND_CLOCE	P25472 clostridium
2	598	34.5	900	1 GUND_CLOTM	P16218 clostridium
3	397	22.9	814	1 GUND_CLOTM	P10477 clostridium
4	370.5	21.4	440	1 GUND_CLOCL	P28621 clostridium
5	356.5	20.6	364	1 GUND_RUMAL	P23660 ruminocecu
6	341.5	19.7	658	1 GUND_FIBSU	P14250 fibrobacter
7	338.5	19.5	515	1 GUND_CLOCE	P28623 clostridium
8	327	18.9	343	1 GUND_CLOCE	P23340 clostridium
9	324	18.7	343	1 GUND_CLOTM	P23661 ruminocecu
10	311.5	18.0	409	1 GUND_RUMAL	P23661 ruminocecu
11	311.5	18.0	473	1 GUND_NEOPA	P23661 ruminocecu
12	311	18.0	517	1 GUND_CLOLO	P54937 clostridium
13	309	17.8	406	1 GUND_RUMAL	P16216 ruminocecu
14	285.5	16.5	475	1 GUND_CLOCE	P17901 clostridium
15	278	16.1	356	1 GUND_RUMFL	P16169 ruminocecu
16	265.5	15.3	366	1 GUND_PAEIA	P23550 paenibacilli
17	233.5	13.5	547	1 GUND_BUTFI	P20847 butyrivibri
18	201.5	11.6	397	1 GUND_PAEPO	P23548 paenibacilli
19	181	10.5	409	1 GUND_BACS4	P06565 bacillus sp
20	180	10.4	488	1 GUND_BACS4	P06565 bacillus sp
21	168	9.7	400	1 GUND_BACS4	P08546 bacillus ag
22	165.5	9.6	1039	1 GUND_CALSA	P10474 c endogluc
23	164.5	9.5	429	1 GUND_BUTFI	P22541 butyrivibri
24	162.5	9.4	448	1 GUND_CLOSA	P15704 clostridium
25	157.5	9.1	499	1 GUND_BACS4	P07983 bacillus su
26	156.5	9.0	499	1 GUND_BACS4	P10475 bacillus su
27	153	8.8	424	1 GUND_RALSO	P58599 ralsstonia su
28	151.5	8.7	499	1 GUND_BACS4	P23549 bacillus su
29	149	8.6	563	1 GUND_CLOTM	P04956 clostridium
30	147.5	8.5	444	1 GUND_ERWCA	P05939 erwina car
31	146.5	8.5	504	1 GUND_ERWCA	P05939 erwina car
32	146.5	8.5	505	1 GUND_ERWCA	P04706 erwina car
33	146	8.4	426	1 GUND_RALSO	P17974 ralsstonia s

34	144.5	8.3	466	1 GUND_THEFU	G01786 thermomonos
35	143.5	8.3	445	1 SPRI_YEAST	P32603 saccharomyc
36	140.5	8.1	459	1 GUND_STRELI	P27035 streptomyce
37	139.5	8.1	418	1 GUND_TREFE	P07982 trichoderma
38	137.5	7.9	484	1 GUND_XANCP	P19487 xanthomonas
39	136.5	7.9	341	1 GUND_CRYFL	P04469 cryptococcu
40	134.5	7.8	825	1 GUND_BACS4	P19570 bacillus sp
41	134	7.7	941	1 GUND_BACS6	P19424 bacillus sp
42	133.5	7.7	566	1 GUND_CLOTM	P00332 clostridium
43	129	7.4	1331	1 MANB_CALSA	P22533 caldocellum
44	124	7.2	651	1 BGLR_HUMAN	P08236 homo sapien
45	121	7.0	570	1 EXGH_SCHPO	P10444 schizosach

ALIGNMENTS

```

RESULT 1
GUND_CLOCE STANDARD; PRT; 584 AA.
ID GUND_CLOCE
AC P25472;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase D)
DE (cellulase D) (EGCD).
GN CELCDD.
OS Clostridium cellulolyticum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1521;
RN [1]
RP SEQUENCE FROM N.A.
RC SRRAT-ATCC 35319;
RX MEDLINE=92009193; PubMed=1916275;
RA Shima S., Igarashi Y., Kodama T.;
RT "Nucleotide sequence analysis of the endoglucanase-encoding gene,
RT celcdd, of Clostridium cellulolyticum."
RL Gene 104:33-38(1991).
CC
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLOBIHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE ST COMPONENT.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSTYL
CC HYDROLASES).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D90341; BAA14354.1;
CC PIR: J01229; J01229;
CC HSSP: P17901; IEDG;
CC InterPro: IPR005087; CBM_11.
CC InterPro: IPR002105; Dockerin_1.
CC InterPro: IPR002048; EF-hand.
CC InterPro: IPR001547; GH_5.
CC Pfam: PF00150; cellulase; 1.
CC Pfam: PF00404; Dockerin_1; 2.
CC Pfam: PF03425; CBM_11; 1.

```

DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00448; CLOS_CELLULOSE_RPT; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 584 CATALYTIC (BY SIMILARITY).
FT DOMAIN 25 328 CATALYTIC (BY SIMILARITY).
FT DOMAIN 329 353 PRO/THR-RICH (LINKER).
FT DOMAIN 354 584 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 530 584 2 x 24 AA APPROXIMATE REPEATS.
FT REPEAT 530 552 1.
FT REPEAT 562 584 2.
FT REPEAT 562 584 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 159 159 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 264 264 OFC41257E81322C3 CRC64;
SQ SEQUENCE 584 AA; 66061 MW; 0FC41257E81322C3 CRC64;
Query Match 34.5%; Score 598; DB 1; Length 584;
Best Local Similarity 37.4%; Pred. No. 4.5e-38;
Matches 116; Conservative 58; Mismatches 114; Indels 22; Gaps 6;
OY 10 KILGRGINIGNALPNEGDMGVIKDEFDILKEAGFSHVRIPIRMSTHVAIVPPYKIM 69
DB 33 KMGIGMNLGNFEDAPTEGWSKAAQYEFDFKQAGFKHVRIPIMDQHTLANSFYTDV 92
OY 70 DFFPKRVDEVINGALKRGGLAVAINIHHEELMNDPEHKERFLAKOLADRYKDPETL 129
DB 93 SNFLNLEIVIMSLSRGVYVINSHDWLMNYSQNGRFEKTIWDLQAKRKSENL 152
OY 130 FFEILNEPKNLTPERKNELLEALVKVIRSIDRKHIIIGTAEWGISALEKLSVPRMEK 189
DB 153 VEIILNEPKNITDSQINDMKNRILNIRKTNTRVVIIGAGVWNSYNSLSQLEIRP-DP 211
OY 190 NSLVTHVYVPEFTTHOGAEWEGSEKWKGRKSGPDDOKHLIEFNFTENSKKKRPI 249
DB 212 NLTATFHYDYPSFTHQ-----WQSGTWGTRKNDMDALIMVFNHWKSDKNIPV 260
OY 250 YIGSEFAYRKADLESRIKMTSFVYREMERKRWMAVWEFCSG-FG-----VYDPLRKT 301
DB 261 YIGSEYGVMGSHDRTSAVKMFEDVSDQKISGFCGAND--NGVFGSDVNDMAYINRDTRO 318
OY 302 WNRDLLEAL 311
DB 319 FDKETLMAIL 328
RESULT 2
GUNH_CLOTM STANDARD: PRT; 900 AA.
ID GUNH_CLOTM
AC P16218;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Endoglucanase H precursor (EC 3.2.1.4) (EGH) (Endo-1,4-beta-glucanase) (Cellulase H)
GN CELH.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9033606; Pubmed=2197182;
RA Yagoue E., Beguin P., Aubert J.-P.;
RT Nucleotide sequence and deletion analysis of the cellulase-encoding gene celH of Clostridium thermocellum.
RL Gene 89:61-67(1990).
CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-GLUCANS.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS

CC WELL AS IN OTHER C. THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SI COMPONENT.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULOSE FAMILY
CC A (FAMILY 5 OF GLYCOSYL_HYDROLASES).
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 26 OF
CC GLYCOSYL_HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M31903; AAA23225.1; -
DR PIR: JH0157; JH0157.
DR HSP: P07985; ICBC.
DR Interpro: IPR005087; CBM_11.
DR Interpro: IPR002105; Dockerin_1.
DR Interpro: IPR002048; EF-hand.
DR Interpro: IPR001547; GH_5.
DR Interpro: IPR00150; Cellulase; 1.
DR Pfam: PF00404; Dockerin_1; 2.
DR Pfam: PF03425; CBM_11; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 44
FT CHAIN 45 900
FT DOMAIN 45 630 CATALYTIC (BY SIMILARITY).
FT DOMAIN 631 654 PRO/THR-RICH (LINKER).
FT DOMAIN 655 900 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 460 460 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 565 565 NUCLEOPHILE (BY SIMILARITY).
FT DOMAIN 833 895 2 x 24 AA APPROXIMATE REPEATS.
FT REPEAT 833 856 1.
FT REPEAT 872 895 2.
SQ SEQUENCE 900 AA; 102415 MW; 973AFB1954FC246B CRC64;
Query Match 34.5%; Score 598; DB 1; Length 900;
Best Local Similarity 36.6%; Pred. No. 7.6e-38;
Matches 115; Conservative 60; Mismatches 121; Indels 18; Gaps 4;
OY 3 VDFPERKNIIGRGINIGNALPNEGDMGVIKDEFDILKEAGFSHVRIPIRMSTHVA 62
DB 327 VDFPERKMKMGKNTNIGNTLEAVYEGWSKSMAYEFDFKQAGKVRIPYRMNHTMR 386
OY 63 FPPYKIMDFEKKRVDEVINGALKRGGLAVAINIHHEELMNDPEHKERFLAKOLADRY 122
DB 387 TYPYTIKAFLDVPEVWVMSLSRGVYVINSHDWIKEDYNGNTEREKTIWDLQAKRK 446
OY 123 KQYPELTFEILNEPKNLTPERKNELLEALVKVIRSIDRKHIIIGTAEWGISALEK 182
DB 447 KNSSENLEFELMNEPKNITDSQINDMKNRILNIRKTNTRVVIIGAGVWNSYNSLS 506
OY 183 SVPRKENSIVTTHVYVPEFTTHOGAEWEGSEKWKGRKSGPDDOKHLIEFNFTENSK 242
DB 507 KIPD-DVYLCTFHYDYPSFTHQ-----WQSGTWGTRKNDMDALIMVFNHWKSD 554
OY 243 KKKRPIYIEFGAYRKADLESRIKMTSFVYREMERKRWMAVWEFCSGGVYD 296
DB 555 DRNNIPYVFEFAYRMAYADRTSRVKWYDFISDALEGRACSVWDNGVFGSLDNDMAIYN 614
OY 297 TLRKTNKDLLEAL 310
DB 615 RDRTFEFLMAIL 628
RESULT 3
GUNH_CLOTM STANDARD: PRT; 814 AA.
ID GUNH_CLOTM

```

AC P10477.1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUL-1989 (Rel. 11, Last annotation update)
DE 01-FEB-1995 (Rel. 11, Last annotation update)
DE Endoglucanase E precursor (EC 3.2.1.4) (EGE) (Endo-1,4-beta-glucanase)
DE (Cellulase E).
GN CELE.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_taxid=1515; [1]
RX SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.
RA MEDLINE=69137992; PubMed=3066698;
RA Hall J., Hazlewood G.P., Barker P.J., Gilbert H.J.;
RT "Conserved reiterated domains in Clostridium thermocellum
RT endoglucanases are not essential for catalytic activity.";
RL Gene 69:29-38(1988).
CC -I- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -I- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCCELLUM CELLULOSONE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -I- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M22759; AAA23224.1; -.
DR PIR; J03447; CZCLEM.
DR HSSP; P17901; IBDG.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR001547; GH_5.
DR InterPro; IPR001087; Lipase_GDSL.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00404; Dockerin_1; 2.
DR Pfam; PF00657; Lipase_GDSL; 1.
DR PROSITE; PS00448; CLOS_CELLULOSONE_RPT; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
FT SIGNAL 1 34
FT CHAIN 35 814 ENDOGLUCANASE E.
FT ACT_SITE 193 193 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 316 316 NUCEOPHILE (BY SIMILARITY).
FT DOMAIN 415 474 2 x 24 AA APPROXIMATE REPEATS.
FT REPEAT 415 438 1.
FT REPEAT 451 474 2.
SQ SEQUENCE 814 AA; 90244 MW; CGFA24B8D1523632 CRC64;
Query March 22.9%; Score 397; DB 1; Length 814;
Best Local Similarity 31.6%; Pred. No. 1.2e-22;
Matches 109; Conservative 57; Mismatches 135; Indels 44; Gaps 11;
QY 3 VDPERNKILRGINIGNALEAPNEDGWG-VVIRKDEFDIKEAGFSHVRIPIRWSTHAY 61
DY :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 58 IDLVKEIK---GMWLGNTLDAPETAMGNPRTTKAMIEKYEMGFNVAVRPVTWDTHIG 114
QY 62 AAPPKKINDRFKKRDEVINGALKRGLAVAINHHYEELM-----NDPEHKERPLALKQ 117
DY :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 115 PAPPYKIDEAWLNREEVVNYVLDCGMVAIIHLHDNTMIIPTYANEOBSKEKLKVWEO 174
QY 118 IADRKQDPPELLFFELINEPHNGNLTPKWK-----NELLEAKVKIRSI----DK 162
DY ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 175 IATFKKDIDLHLETETNMREPREVGSPMEMMGCTYENRDVINRFNLAAYNTITRASCGNNCK 234

```

ID	NAME	FUNCTION	ORGANISM	SEQUENCE
0y	163	KRTII1GAEWGQISALAEKLSKPKKSKSIYTIHYVNFPEETHOCAEVESEKWLGRK	222	
Db	235	REL1VPTNNAAGLDVALNDVIYPNNDKRVYISIHAYSYTF-----ADVNQTS-----YW	285	
0y	223	GSPDDKHLIEEFNFIEEMSKNKRPIYIGFGAYRKADLESRIKMTSFVREMEKRRWS	282	
Db	286	GSDYKRASTLSTSLDAIYRNFYVNGRAVYIIGEGFTIDKNNLSRYAAEHYRAEVSRCIA	345	
0y	263	WAYMEFCGF---GYVDL---RKT---WNKDLLEALIGDSIE	317	
Db	346	VFWWD-NGYVNPDAETVALNRTLTSMVYPEIYQALMRGAVE	388	

RESULT 4

ID	GUNB_CLOCL	STANDARD	PRT	440 AA
AC	P28621			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-D-glucanase B) (Cellulase B).			
GN	ENGB.			
OS	Clostridium cellulovorans.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.			
OX	NCBI_TaxID=1493;			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 35296;			
RX	MEDLINE=92065240; PubMed=1955860;			
RT	Foong F., Hamamoto T., Shoseyov O., Doi R.H.:			
RT	"Nucleotide sequence and characteristics of endoglucanase gene engB from Clostridium cellulovorans."			
RL	J. Gen. Microbiol. 137:1729-1736(1991).			
CC	-1- FUNCTION: HAS ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYL-CELLULOSE (CMC), XYLAN AND LICHENAN, BUT NOT AVELL.			
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.			
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic linkages in xylans.			
CC	-1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.			
CC	-1- MISCELLANEOUS: UP TO 32 AA OF THE N-TERMINUS AND 52 AA OF THE C-TERMINUS ARE NOT REQUIRED FOR CATALYTIC ACTIVITY.			
CC	-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).			
CC	-----			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION			
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - LIS			
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS			
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY			
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL			
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/or-send-an-email-to-licence@isb-sib.ch).			
CC	-----			
DR	EMBL: M75706; AAA23231.1; -			
DR	HSSP: P17901; 1EDG.			
DR	InterPro: IPR002105; Dockerin_1.			
DR	InterPro: IPR001547; GH_5.			
DR	Pfam: PF001150; cellulase; 1.			
DR	Pfam: PF00404; Dockerin_1; 2.			
DR	PROSITE: PS00018; EF_HAND; UNKNOWN_2.			
DR	PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.			
DR	PROSITE: PS00659; GLYCOSYL_HYDROL_F3; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.			
FT	CHAIN	1	33	POTENTIAL.
FT	CHAIN	34	440	ENDOGLUCANASE B.
FT	ACT_SITE	179	179	PROTON DONOR (BY SIMILARITY).
FT	ACT_SITE	305	305	NUCLEOPHILE (BY SIMILARITY).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M29047; AAA24893.1; -
DR PIR: A33598; A33598.
DR HSP: P07985; 1CE0.
DR InterPro: IPR005087; CBM_11.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF03425; CBM_11; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Lipoprotein;
KW Signal.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 265
FT CHAIN 266 658
FT LIPID 24 24
FT ACT_SITE 448 448 N-ACTY DIGLYCERIDE (POTENTIAL).
FT ACT_SITE 597 597 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 658 AA; 73424 MW; 1C96E64C3F7109A6 CRC64;

Query Match 19.7%; Score 341.5; DB 1; Length 658;
Best Local Similarity 26.4%; Pred. No. 1.5e-18;
Matches 92; Conservative 57; Mismatches 145; Indels 55; Gaps 12;

OY 10 KIIRGINIGNALEAENEGDVGVIKDEFEEDIKAGFSHVRIPI---RMSTHAYAF--- 63
DB 308 KYLNKGVNTNMLENDGKSEFELGSDVKILADNGFSLRPIDDLATNRDAFIAG 367
OY 64 --PPIYI-MDREFKRVDEYINGALRGALAVAINIHYEELMNDPEEKERETAL---WK 116
DB 368 TDTLKFDDDTLEFLVDSFEVETAKYMSFVIDYHEYDMSYNTSAKDPYIKMAETWK 427
OY 117 QIADRKQYR-ETLFEELINER---HGNTLPEKMNLEELALKVTSIDKHTIIIGTA 172
DB 428 HVAHAHAEPSREDLEFELNPEPMSDGKVAATWTTPAAGAMIDAIKTVTKHTILFGDAQ 487
OY 173 WGISALEKLSVPKMEKNSIVTTHYVNPFEETHOGAEWVE-----GSEK--- 217
DB 488 KWSITLLAK-RIPFTDNNIIYVHTPEAFTHOGGSMTDYATIHIDIPPYDAKMSYVS 546
OY 218 --LG-----RKWSPDQKHLEEFNLEIEMSKKKRPYIGEGAYR-KAD 261
DB 547 GDFGVKSKTSYVKTNIKKNYKTSKEALIEQLKAKKMAATNNVPIINIEFGALMLRST 606
OY 262 LESRIKTSFYVAREMKRMSAYMEFCGFGYDRLRTKWNDDLEAL 310
DB 607 AESRLNLTAMREICDTLOIPWTHMGYIGNFSYIE-----NGKLIEGL 649

RESULT 7
GUND_CLOCL STANDARD: PRT: 515 AA.
AC P28623;
DT 01-DEC-1992 (rel. 24, Created)
DT 01-DEC-1992 (rel. 24, Last sequence update)
DT 01-JUN-1994 (rel. 29, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase C)
DE (Cellulase D).
GN ENCD.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1493;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-44.

RC STRAIN-ATCC 35296;
RA MEDLINE=92167968; PubMed=1538700;
RX Hamamoto T., Poong F., Shoseyov O., Dol R.H.;
RT "Analysis of functional domains of endoglucanases from Clostridium
RT cellulovorans by gene cloning, nucleotide sequencing and chimeric
RT protein construction.";
RL Mol. Gen. Genet. 231:472-479(1992).
CC -1- FUNCTION: HAS ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYL-CELLULOSE
CC (CMC), CELLULOBIOSIDASE ACTIVITY ON P-NITROPHENYL-CELLULOSE
CC (P-NPC), AND PARTIAL HYDROLYTIC ACTIVITY ON CRYSTALLINE CELLULOSE
CC (AVICEL).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M37434; AAA23233.1; -
DR HSP: P07986; 1EXG.
DR InterPro: IPR001919; Bac_Cellose-bind.
DR InterPro: IPR001547; GH_5.
DR InterPro: IPR001230; Prenyl_site.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00553; CBM_2; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
KW SIGNAL 1 31
FT CHAIN 32 515
FT DOMAIN 32 376
FT DOMAIN 377 407
FT DOMAIN 408 515
FT ACT_SITE 180 180 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 303 303 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 515 AA; 55976 MW; 4CEB736CE76373F0 CRC64;

Query Match 19.5%; Score 338.5; DB 1; Length 515;
Best Local Similarity 27.6%; Pred. No. 1.9e-18;
Matches 92; Conservative 62; Mismatches 138; Indels 41; Gaps 11;

OY 15 GINIGNALEA-PNEGDMG-VVINKDEFEEDIKAGFSHVRIPIRMSTHAYAFPPYKIMDRF 72
DB 53 GWNLTGMTAIDGGETMNGMNTNHAMINKIEAGFMTLRPVWDGIMGAPEYTDIDOTW 112
OY 73 FKRVDEYINGALRGALAVAINIHYEELAN---DDEHKERLALMKQIADRKQYR 128
DB 113 MKRVEEIANAFNDWYVILNLHNEHEDWKPFYANAQVAAQLTKWTOIANNFKKGGDH 172
OY 129 LFEELINEPHGNLTPEK---NELLEALKVTRSIDKHT---IITGAEMG 174
DB 173 LIFETMNERPRPVASLQWIGSGSYENREYVNRKYLTAVNAIRAGGNATRYINVPILAS 232
OY 175 GIS-ALEKLSVPKMEKNSIVTTHYVNPFEETHOGAEWVEGSEKWLGRKWSDDOKHLE 233
DB 233 AMSTTINDVILIPNDSKVIYLSHMSYPF---AMDINGTS-----SMGSDYKSLDS 283
OY 234 EFNFIEMSKKKRPYIGEGFAYRKADLESRIKWSFYVAREMKRMSAYME----- 287
DB 284 EFDAYVYKPFVKNRAVAVIGMGISINKNNTAARVTHAEVYAKSAKARGLPIWMDNGSYA 343
OY 288 -FCSGGVYDTLRKTNW-KDLLEAL---IGGDS 315
DB 344 GRAFTGIGFIRNSLUTWDAPVEMKAFIKIGGSS 376


```

RESULT 8
GUNC_CLOSF STANDARD: PRT: 343 AA.
ID GUNC_CLOSF
AC P23340;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase C307 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN CELC307.
OS Clostridium sp. (strain F1).
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1508;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-19.
RX MEDLINE=91299276; PubMed=1368690;
RA Sakka K., Shimanuki T., Shimada K.;
RT "Nucleotide sequence of celc307 encoding endoglucanase C307 of
RT Clostridium sp. strain F1."
RL Agric. Biol. Chem. 55:347-350(1991).
CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- PTM: THE SIGNAL PEPTIDE WAS NOT CLEAVED IN THE PROTEIN EXPRESSED
CC IN E.COLI.
CC -1- SIMILARITY: IDENTICAL TO THE ECG FROM C.THERMOCELLUM, EXCEPT FOR
CC THREE POSITIONS.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D00945; BAA00793.1; -
CC PIR: J04009; J04009.
CC DR HSSP: P07985; ICEN.
CC DR InterPro: IPR001347; GH_5.
CC DR Pfam: PF00150; cellulase; 1.
CC DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC KM Cellulose degradation; Hydrolase; Glycosidase; signal.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 343 ENDOGLUCANASE C307.
CC FT ACT_SITE 140 140 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 280 280 NUCLEOPHILE (BY SIMILARITY).
CC FT ACT_SITE 280 280
CC SQ SEQUENCE 343 AA; 40905 MW; 72DD9BEA01ADD05 CRC64;
CC -----
Query Match 18.9%; Score 327; DB 1; Length 343;
Best Local Similarity 28.1%; Pred. No. 8.6e-18;
Matches 97; Conservative 61; Mismatches 125; Indels 62; Gaps 13;

```

```

DB 176 YNSDELKNTLA--DIDDYIYNNFHYNPEFTHQKAMHSESAAMAYNRKYPGQYEGIE 233
QY 226 DDQKH-----LIEPNFIE-----EWSKKNRPYIIGEPGAYRRADLESRI 266
DB 234 EFVKNNPKYSFPMELNNLKLNRKDLPAIDFERKKCKLYCGEYAIADLESRI 293
QY 267 KWTSTVREMEKRRMSMAVMEFCS-GFGVDTLTKRTNNKLLLEAL 310
DB 294 KWHEDIISLEBYDIGAVWNYKKMDFEITYNEDKRPVSGELVNTL 338

RESULT 9
GUNC_CLOTM STANDARD: PRT: 343 AA.
ID GUNC_CLOTM
AC P07985;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endoglucanase C (EC 3.2.1.4) (EcC) (Endo-1,4-beta-glucanase)
DE (Cellulase C).
GN EcC.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88255856; PubMed=3384335;
RA Schwarz W.H., Schlumming S., Ruecknagel K.P., Burgschwalger S.,
RA Krell G., Staudenbauer W.L.;
RT "Nucleotide sequence of the celC gene encoding endoglucanase C of
RT Clostridium thermocellum."
RL J. Biol. Chem. 268:14096-14102(1993).
RN [2]
RP ACTIVE SITE GLU-280.
RX MEDLINE=93300798; PubMed=8100226;
RA Wang Q., Tull D., Melnke A., Gilkes N.R., Warren R.A., Aebetsold R.,
RA Withers S.G.;
RT "Glu280 is the nucleophile in the active site of Clostridium
RT thermocellum CelC, a family A endo-beta-1,4-glucanase."
RL J. Biol. Chem. 268:14096-14102(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).
RC STRAIN=NCIB 10682;
RX MEDLINE=95393242; PubMed=7664125;
RA Dominguez R., Souchon H., Spinnell S., Dauter Z., Wilson K.S.,
RA Chauvaux S., Beguin P., Alzari P.M.;
RT "A common protein fold and similar active site in two distinct
RT families of beta-glucanases."
RL Nat. Struct. Biol. 2:569-576(1995).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF MUTANT GLN-140.
RC STRAIN=NCIB 10682;
RX MEDLINE=96192088; PubMed=8632467;
RA Dominguez R., Souchon H., Lascombe M.-B., Alzari P.M.;
RT "The crystal structure of a family 5 endoglucanase mutant in
RT complexed and uncomplexed forms reveals an induced fit activation
RT mechanism."
RL J. Mol. Biol. 257:1042-1051(1996).
CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- PATHWAY: Cellulose degradation.
CC -1- MISCELLANEOUS: EcC HAS AN UNUSUAL SUBSTRATE RANGE AND DISPLAYS
CC FEATURES COMMON TO CELLIOHYDROLASES BY BEING ABLE TO CLEAVE THE
CC ACIDIC BOND OF ARYL-BETA-GLUCOSIDES.
CC -1- SIMILARITY: IDENTICAL TO THE ECG FROM CLOSTRIDIUM SP. STRAIN F1,
CC EXCEPT FOR THREE POSITIONS.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M19422; AAA3220.1; -
DR PIR: J02668; J02668.
DR PDB: 1CEN; 03-APR-96.
DR PDB: 1CE0; 03-APR-96.
DR PDB: 1CEC; 29-JAN-96.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase.1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
FT ACT_SITE 140 140
FT ACT_SITE 280 280 NUCLEOPHILE.
SQ SEQUENCE 343 AA; 40954 MW; BD24763F548C726E CRC64;

Query Match 18.7%; Score 324; DB 1; Length 343;
Best Local Similarity 28.1%; Pred. No. 1.5e-17;
Matches 97; Conservative 61; Mismatches 125; Indels 62; Gaps 13;

OY 15 GING---NALAPNEGDMGVIRKDEFFDIKEAGSFHVRIPRSTHAYAFPPYKIMDR 71
DB 7 GINLGMISQYQFSKHFPTTEKDIETIARGFPHVLP-----FDPIIESDDN 59
OY 72 F-----FRVDEVINGALRGLAVAINIH-----YEELNDPEHEKRPAL 114
DB 60 VGEYKEDGLSYDRCLMECKKYNGLVDMHNAQYRFQDFKSTLFEDPNQOK-REVDI 118
OY 115 WKIADRYKRYPTLPPELLNEPHGNLTPE--KNELLEALAVIRIDKKHIIITAE 172
DB 119 WRFLARKYINERHIAELLNEV---VEPSTRNKLMLTYIKAIRIDSTMLTYIGNN 175
OY 173 WGISALEKLSVPRKENSIV-TIHYYNPEETHOGAEWEG-----SEKWLGRKMGSP 225
DB 176 YNSPDELKNLA--DIDDDYIVNPHFYNPFFTHQKAMHSESAMAYRYKYRQYIGIE 233
OY 226 DDOKH-----LIEEFNIE-----EWSKKNRPIYIGEGAYRRADLESRI 266
DB 234 EFKYNNKYSFEMLENNLKLKELLRKDLPAIEFKEKCKKLYCGEFGVIAIADLESRI 293
OY 267 KMTSFVYREMEKRRMSWAYWFCSS-GFGVYDTLRKTNKLLLEAL 310
DB 294 KMHEDIYISLEEDYDGGAVNNYKMDPEYINEDKRPVSOELVNTL 338

RESULT 10
GUNB_RUMAL STANDARD; PRT; 409 AA.
AC P23661;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (ECB).
GN CELB.
OS Ruminococcus albus.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
OC Ruminococcus.
OX NCBI_TaxID=1264;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SY3;
RX MEDLINE=91066833; PubMed=2250649;
RA Poole P.M., Hazlewood G.P., Laurie J.L., Barker P.J., Gilbert H.J.,
RT "Nucleotide sequence of the Ruminococcus albus SY3 endoglucanase
RL genes cels and celsB."
CC Mol. Gen. Genet. 223:217-223(1990).
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic

CC linkages in cellulose.
CC -I- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X54932; CAA38693.1; -
DR PIR: S12018; S12018.
DR HSSP: P17901; 1EDG.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase.1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; UNKNOWN_1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; signal.
FT SIGNAL 1 21
FT CHAIN 22 409
FT ACT_SITE 212 212
FT ACT_SITE 332 332 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 409 AA; 45523 MW; 119081DFA3BFDD54 CRC64;

Query Match 18.0%; Score 311.5; DB 1; Length 409;
Best Local Similarity 27.2%; Pred. No. 1.0e-16;
Matches 88; Conservative 51; Mismatches 139; Indels 45; Gaps 10;

-OY 12 IGRGINGNLAPNEN-----DW-----GVYKDEFFDIKEAGSHRIPRSTHAYAF 63
DB 78 MSNOMNLGNTMDATGEBLESEISLPTKYTNKFMIDMLPAGNNVLRIPVSNHLID- 136
OY 64 PPKYKMDRFRKRVDEVINGALRGLAVAINIHVEELNDPEEHK---ERFLAKIOIAD 120
DB 137 NNYTIDPAMMDRVOEIVNYGIDDDGMVILNTHHEEWYMPKREKDGIEELKAIMSOIAD 196
OY 121 RYKDYPTLPPELLNEPHGNLTPEKNELLEALVKYRSIDK-----KHT 165
DB 197 RFKGYDHLTFEGINERLRNGEGAEWGTG-TSEARELINNEYKAEVETVRASGGNGRCL 255
OY 166 IITGAEWGISALEKLSVPRKENSIVTIHYYNPEETHOGAEWEGSEKWLGRKMGSP 225
DB 256 MITGYAASGVNNLSALELDEDSKLLISVHAYLPYSF-----ALDTGTGTXY-----DP 305
OY 226 DDOKHLLIEFNFIEMSKKRPYIYIGEGAYRRADLESRIKWTSFVYREMEKRRMSWAY 285
DB 306 EDTA-IPTLFESLNEFLISRDIPVIVGEFGSMNKNIDRVCIDYDLGNAAKYDIPCVW 364
OY 286 WEFCSS-----GFGVYDTLRKTTW 302
DB 365 WDNVARIENGNEFGLNRQEDW 387

RESULT 11
GUNB_NEOPA STANDARD; PRT; 473 AA.
AC Q12647;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase B)
DE (Cellulase B).
GN CELB.
OS Neocallimastix patriciarum (Rumen fungus).
OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
OC Neocallimastixaceae; Neocallimastix.
OX NCBI_TaxID=4758;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94128068; PubMed=8297343;
RX Zhou L., Xue G., Orpin C.G., Black G.W., Gilbert H.J., Hazlewood G.P.;

"Ironless celb from the anaerobic fungus Neocallimastix patriciarum RT encodes a modular family A endoglucanase."

RL Biochem. J. 297:359-364(1994).

CC -1- FUNCTION: RATE OF HYDROLYSIS OF CELLULO-OLIGOSACCHARIDES INCREASED WITH INCREASING CHAIN LENGTH FROM CELLULOSE TO CELLOPENTAOSE.

CC -1- CATALYTIC ACTIVITY: Endohydrolase of 1,4-beta-D-glucosidic linkages in cellulose.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: 231364; CAA83238.1; .

DR HSSP: P17901; IEDC.

DR InterPro: IPR002883; CBD_5.

DR InterPro: IPR001547; GH_5.

DR InterPro: IPR001230; Prenyl_site.

DR InterPro: IPR001202; WM_hsp_mmp.

DR Pfam: PF00150; cellulase; 1.

DR Pfam: PF02013; CBM_10; 2.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

DR PROSITE: PS01159; WM_DOMAIN_1; 1.

DR Hydrolase; Glycosidase; Repeat; Signal.

FT SIGNAL 1 17

FT CHAIN 18 473

FT DOMAIN 18 367

FT DOMAIN 367 387

FT DOMAIN 387 376

FT DOMAIN 376 381

FT DOMAIN 390 473

FT REPEAT 390 427

FT REPEAT 435 473

FT ACT_SITE 173 173

FT ACT_SITE 295 295

FT ACT_SITE 473 AA; 53070 MW; B1ID9F171EA33199 CRC64;

SO SEQUENCE

Query Match 18.0%; Score 311.5; DB 1; Length 473;

Best Local Similarity 28.0%; Pred. No. 1.9e-16;

Matches 98; Conservative 53; Mismatches 136; Indels 63; Gaps 13;

10 KILRGINIGALAE-----PNEGDMG-VYIKDFDIKEAGFSHVRIPIR 55

30 KDLITGSLGTLDTATCFETLDYKKNQIASETCMGNNKTTQELYYKLSDLGFTFRITPT 89

56 WSTHAAYAPPYKINDREFKRVDEVINGALRGIAVAIINH--YEELMDPEHKEKFL 112

90 WSGHFGNAPDKINDQWNRVHEIYDAINTGYALINHEHETNNAHQKMLSEAKKLLV 149

113 ALMKOIAIDRYDYPTLFEFLINPEHGNLTPPEK-----NELEALKVIRSID 161

150 AIWKOIAIEFADYDEHLLFEENNEPRKVGDPRENNNGDYEENNEVNDLFEVKTIRATG 209

162 -----KKHTIITIGTAEWGISALEKLVPRKEKNSIVTTHYNNPEFT-HQGA---EWVE 212

210 GNNALRHLMIPYACINDGAIINNKKPPSGDKVIVLSHSVYFALNNGAISAISNYD 269

213 GSE-KWIGRKWSPDDQKHLEEFNFIEWSEKKNRPYIIGFGAYRKADLESRIKWTSE 271

270 GSEIWM-----AMNTINSKPIISRGIPVIGEGAMNRNEDDEREMAEY 313

314 VYREMEKRRWSWAYME--FCSG---FGYD--TLRKTNKKLLBALIGG 313

314 YIKKATISGVPCVIMDNGIFEGEGERFGLINSTIQVYVPR-LVNGLLKG 362

RESULT 12

GUNA_CIOLO

ID GUNA_CIOLO STANDARD: PRT; 517 AA.

AC P54937;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A) (cellulase A).

GN CELA.

OS Clostridium longisporum.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1523;

RN [1]

RP STRAIN=ATCC 49440;

RC MEDLINE=94172316; PubMed=8126442;

RA Mientdorf V., Thomson J.A.;

RT "Cloning of an endo-(1-->4)-beta-glucanase gene, cels, from the rumen bacterium Clostridium sp. ('C. longisporum') and characterization of its product, cels, in Escherichia coli."

RL J. Gen. Microbiol. 139:3233-3242(1993).

CC -1- FUNCTION: THE PH AND TEMPERATURE OPTIMA OF CELA ARE 4.8 AND 43 DEGREES CELSIUS, RESPECTIVELY. IT HYDROLYSES BARLEY BETA-GLUCAN, LICHENAN, CARBOXYMETHYLCELLULOSE AND XILAN. IT SHOWS PREFERENTIAL ACTIVITY AGAINST THE LARGER CELLOOLIGOSACCHARIDES (CELLOHEXAOSE AND CELLOPENTAPOSE); CELLOTETRAOSE IS THE SMALLEST SUBSTRATE DEGRADED COMPLETELY.

CC -1- CATALYTIC ACTIVITY: Endohydrolase of 1,4-beta-D-glucosidic linkages in cellulose.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).

CC -1- SIMILARITY: CONTRAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: L02868; AAC37035.1; .

DR HSSP: P17901; IEDC.

DR InterPro: IPR001919; Bac_celose-bind.

DR InterPro: IPR001547; GH_5.

DR Pfam: PF00150; cellulase; 1.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

DR Cellulose degradation; Hydrolase; Glycosidase; Signal.

KW SIGNAL 1 25

FT CHAIN 26 517

FT DOMAIN 26 517

FT DOMAIN 421 517

FT ACT_SITE 185 185

FT ACT_SITE 309 309

FT ACT_SITE 517 AA; 57660 MW; A1D1570302FEBA30 CRC64;

SO SEQUENCE

Query Match 18.0%; Score 311; DB 1; Length 517;

Best Local Similarity 30.0%; Pred. No. 2.3e-16;

Matches 103; Conservative 45; Mismatches 143; Indels 52; Gaps 14;

12 LGRGINIGALAE-----PNEGDMG-VYIKDFDIKEAGFSHVRIPIRST 58

46 MGVMGNIENGLDAKITNLVNTSPISFETGWNPVYTKAMIDIKNAGKRTIRIPTWGE 105

59 HAYAFPPYKINDREFKRVDEVINGALRGIAVAIINHVEELINDEPEEKE-----RELA 113

106 HLDG--NNKLNIEWKRVKEVDCIADDIYVILINH--EGNVVITPYAKKESSVPTKLT 162

114 LMKOIAIDRYDYPTLFEFLINPEHGNLTPPEK-----NELEALKVIRSI-- 160

163 LMTQISEAFKDYDHLIFELINPEHGNLTPPEK-----NELEALKVIRSI-- 222

Db	194	DRFGYDEHLIFEBLNPRLRGCAEWTG-TSARELINEYEAFVETYRASGANNDR	252
Qy	165	TIIGTAEWGSISALEKLSVPKMEKNSIVTIIHYNPEFTHOGAEWVEGSEKMLGRKMS	224
Db	253	LMITGVAASSAYNNLSAIELPEPDSDKLIISVHAYLPYSF-----ADLTGCTDKY-----D	302
Qy	225	PDDCKHLEENFNFLEEMSKSKNRPIYIGEGAVRKADLESRIKWIISFVYREMEKRRMSA	284
Db	303	PEPDA-IPLEFHEHNELEFISKGPVIVIGEGFTMNKENTEDRYCLEDYLAAMAKYDPCV	361
Qy	285	YMECS-----GRCGYDTRLKPMN-KDLE	308
Db	362	WMDYVARIENGENFGMLNRADLEWPEPDLIE	392
RESULT 14			
GUNA.CLOCE			
ID	GUNA.CLOCE	STANDARD;	PRT; 475 AA.
AC	P17901;		
DT	01-NOV-1990 (Rel. 16, Created)		
DT	01-NOV-1990 (Rel. 16, Last sequence update)		
DT	01-NOV-1997 (Rel. 35, Last annotation update)		
DE	Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)		
DE	(cellulase A) (EGCCA).		
GN	CEJCCA.		
OS	Clostridium cellulolyticum.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OX	NCBI_TaxID=1521.		
RN	[1]		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 27-39.		
RC	STRAIN=ATCC 35319;		
RX	MEDLINE=90108715; PubMed=2558058;		
RA	Faure E., Belaich A., Bagnara C., Gaudin C., Belaich J.-P.;		
RT	"Sequence analysis of the Clostridium cellulolyticum endoglucanase-A-		
RT	encoding gene, celCCA.";		
RT	Gene 84:39-46(1989).		
RN	[2]		
RP	CHARACTERIZATION.		
RX	MEDLINE=92078105; PubMed=1744052;		
RA	Fierobe H.-P., Gaudin C., Belaich A., Loufti M., Faure E., Bagnara C.		
RT	Baty D., Belaich J.-P.;		
RT	"Characterization of endoglucanase A from Clostridium		
RT	cellulolyticum.";		
RT	J. Bacteriol. 173:7956-7962(1991).		
RN	[3]		
RP	MUTAGENESIS OF ARG-104; HIS-147 AND HIS-148.		
RX	MEDLINE=92325059; PubMed=1624455;		
RA	Belaich A., Fierobe H.-P., Baty D., Busetta B., Bagnara-Tardif C.,		
RA	Gaudin C., Belaich J.-P.;		
RT	"The catalytic domain of endoglucanase A from Clostridium		
RT	cellulolyticum: effects of arginine 79 and histidine 122 mutations on		
RT	catalysis.";		
RT	J. Bacteriol. 174:4677-4682(1992).		
RN	[4]		
RP	X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 27-406.		
RC	STRAIN=ATCC 35319;		
RX	MEDLINE=96097400; PubMed=8555787;		
RA	Ducros V., Czjzek M., Belaich A., Gaudin C., Fierobe H.P.,		
RA	Belaich J.-P., Davies G.J., Haer R.;		
RT	"Crystal structure of the catalytic domain of a bacterial cellulase		
RT	belonging to family 5.";		
RT	Structure 3:939-949(1995).		
CC	-I- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE		
CC	GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:		
CC	(1) ENDOCELLULOANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;		
CC	(2) EXOCELLULOBIODHOLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE		
CC	FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;		
CC	(3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYSE THE CELLOBIOSE AND OTHER		
CC	SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.		
CC	-I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic		
CC	linkages in cellulose.		

```

CC      -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC      WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC      MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT
CC      -1- MISCELLANEOUS: THE C-TERMINUS (AA 411-475) MAY PLAY A ROLE IN
CC      -1- ORGANIZING THE CELLULOSE COMPLEX.
CC      -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC      HYDROLASES).
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb.ch/announcement/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M93096; AAA51444.1; -.
CC      DR EMBL: M32362; AAA3221.1; -.
CC      DR PIR: JH0082; CZCLCA.
CC      DR PDB: 1EDG; 17-AUG-96.
CC      DR InterPro: IPR002105; Dockerin_1.
CC      DR InterPro: IPR002048; EF-hand.
CC      DR InterPro: IPR001547; GH_5.
CC      DR Pfam: PF00150; cellulase.1.
CC      DR Pfam: PF00404; Dockerin_1; 2.
CC      DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
CC      DR PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.
CC      DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC      KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal;
CC      3d-structure.
CC      KM SIGNAL 1 26
CC      FT CHAIN 27 475 ENDOLUCANASE A.
CC      FT ACT_SITE 147 147 PROBABLE.
CC      FT ACT_SITE 195 195 PROTON DONOR.
CC      FT ACT_SITE 332 332 NUCLEOPHILE.
CC      FT DOMAIN 415 469 2 X 24 AA APPROXIMATE REPEATS.
CC      FT REPEAT 415 438 1.
CC      FT REPEAT 446 469 2.
CC      FT MUTAGEN 104 104 R->K: SMALL LOSS OF ACTIVITY.
CC      FT MUTAGEN 104 104 R->S,V: LARGE DECREASE OF ACTIVITY.
CC      FT MUTAGEN 147 147 H->S,E,G,F: TOTAL LOSS OF ACTIVITY.
CC      FT MUTAGEN 148 148 H->V: LARGE DECREASE OF ACTIVITY.
CC      SQ SEQUENCE 475 AA; 53625 MW; 1AF20EA2C0A132F9 CRC64;

Query Match 16.5%; Score 285.5; DB 1; Length 475;
Best Local Similarity 28.4%; Pred. No. 1,8e+14;
Matches 97; Conservative 49; Mismatches 130; Indels 65; Gaps 15

OY 10 KILGRGININALEAPN-----EGDM-GVLIKDEFFDIIRKGFSHVRIPRMST 58
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 52 KGLRLGNLNTDADFNGTMITNELDETSMGSIKTKQMDAIKQGFNTRIPIVSMHP 111
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 59 HAYAFPRYKIMDRPREKRVDEVINGALKRGLAVAINIHH-----YELMNDPEHKEF 111
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 112 HV-SGSYRKISDVWMMNRVQEVVNYNCIDNKMVYVILNTHHDVDKYKGYRPSQYMASSKRYT 170
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 112 LALMKQIADYVKQPELTFEELNEP-----HGNTLPKXWMLL-EALAKYRSIDKKHTI 166
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 171 TSYWAQIAPRANDELHILEGNNEPRVLGHAN---EMWPELTNSDVVDSTINCINOLNOD 227
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 167 IIGTAE-WGGISALEKLSVRKW-----EKSNIYIHYINPEFT 204
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 228 FVNIVRATRGGMASRYLMCPGYVASPDGATNDYFRMENDISGNNNKKTIVSHVACPNMFA 287
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

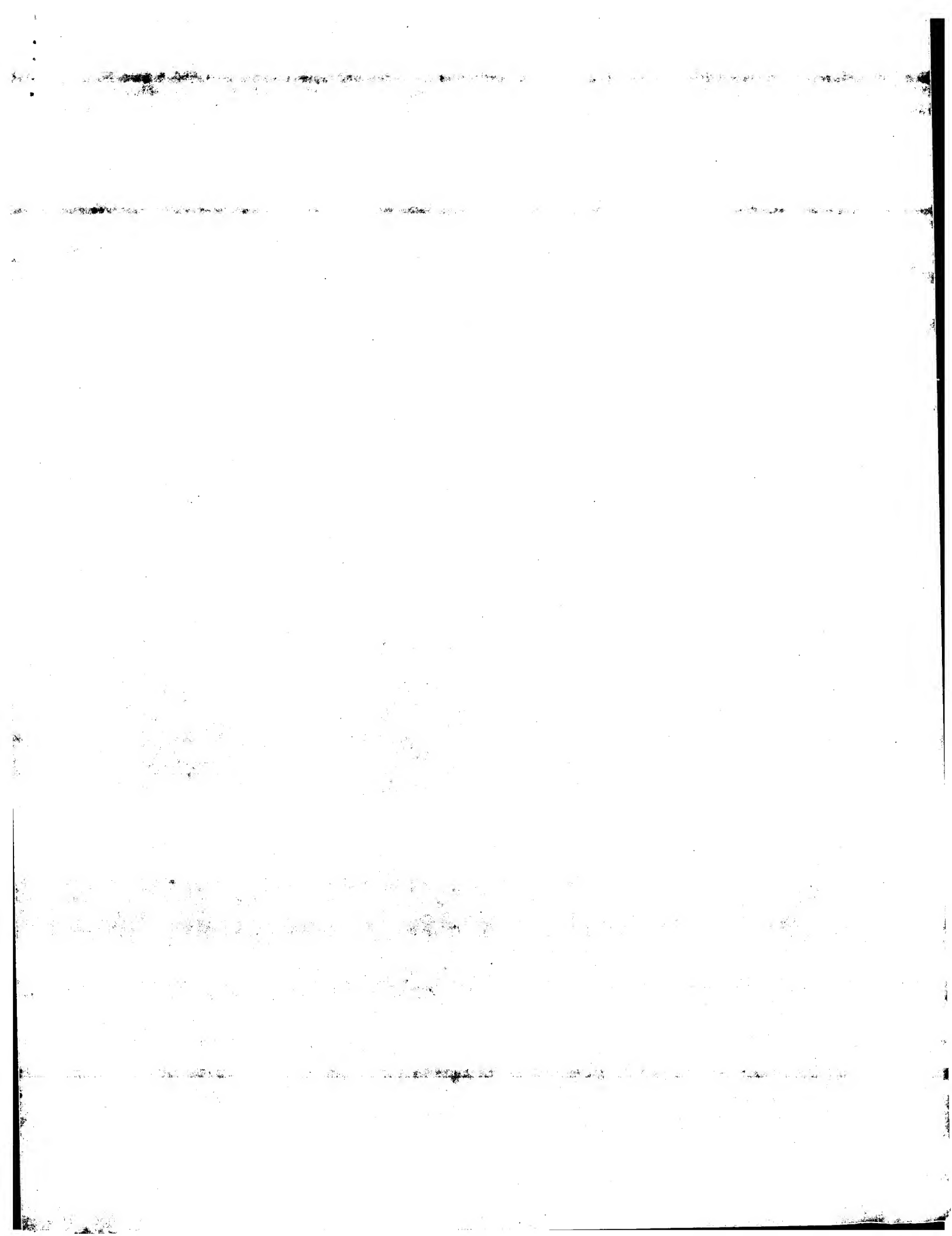
OY 205 HOGAEWEGSEKWLGRKMGSPDDOKHLIEPNFIEE-WSKNKR--PLYGEGAFYRKAD 261
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 288 --GIAMADG-----GTNMMNIDSKDSEVYTWEMDNITNKTSGIPVILIGECAYDKNN 340
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 262 LESRIKWTSEFVREMEKRRMSWAYNE--FCSG-----EGVYD 296
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 341 LKTRVEYMSYVAAQAKARGILCIIMDNNNFSGTSELGCFPD 381
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

ID	NAME	STANDARD	PRT	336 AA
AC	GUIN_RUMFL			
AC	P16169			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	01-FEB-1996 (Rel. 33, Last annotation update)			
DE	Cellodextrinase A (EC 3.2.1.17).			
GN	CELA.			
OS	Ruminococcus flavefaciens.			
OC	Bacteria: Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;			
OC	Ruminococcus			
OX	NCBI_TaxID=1265;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-FD-1;			
RC	MEDLINE=91109710; PubMed=2129544;			
RX	Mang W., Thomson J.A.: Pubmed=2129544;			
RA	"Nucleotide sequence of the <i>celA</i> gene encoding a cellodextrinase of			
RT	Ruminococcus flavefaciens FD-1,"			
RL	Mol. Gen. Genet. 222:265-269(1990).			
RN	(2)			
RP	REVISIONS, SEQUENCE FROM N.A.			
RC	STRAIN-FD-1;			
RL	Thomson J.A.;			
RL	Submitted (Mar-1994) to the EMBL/GenBank/DBJ databases.			
CC	- FUNCTION: CRYSTALLINE CELLULOSE DEGRADATION.			
CC	- SUBCELLULAR LOCATION: SECRETED; SEEMS TO BE EXPORTED BY A			
CC	MECHANISM OTHER THAN A LEADER PEPTIDE.			
CC	- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on lists			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X51944; CA36207.1; -			
DR	PIR; S11926; S11926.			
DR	HSSP; P07985; 1CEO.			
DR	InterPro; IPR001547; GH_5.			
DR	Pfam; PF00150; cellulase; 1.			
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase.			
FT	ACT SITE 141 141			
FT	PROTON DONOR (BY SIMILARITY).			
SO	SEQUENCE 336 AA; 38686 MW; 7B3DB82E3E346711 CRC64;			
Query Match	16.1%; Score 278; DB 1; Length 336;			
Best Local Similarity	26.5%; Pred. 4.4e-14;			
Matches	79; Conservative 55; Mismatches 132; Indels 32; Gaps 8			
QY	14 RGINIGALNAPNEGDMGVYIKDEFF---DI--IKERGFSHVYPIPIRMSTHNAFAFPYKI 68			
DB	9 KGIMLGWM--SQCDYSRERLDNFVKEKDIKQIADWGFPHVRLPIDY--NIVQNDGSV 63			
QY	69 MDRFRKRVDEVINALKRLGLAVAINIH-----HYELMDNDEENKREFLALMKOI 118			
DB	64 IEDGVYINIDIVVELICRRYVGLKLVYIDLKTKTGFSFDGEPSPSGFNDKNEYERLYIMEEL 123			
QY	119 ADKKADPEPTLFEFLINPEHGNLIPKWNLELBALVKVIRSIDKHTIITIGLAEWGISA 178			
DB	124 ARRGCHDTNIVFLLNLEVDENATIGKWNLEISDICIRIKRIAPVEYIITLLGSYHNNAADT 183			
QY	179 LEKLSVPKWEKNSIVTITHYINPEFTTHQGALEWVDSGSKLGRKWSGSDDKHLLDE-FNF 237			
DB	184 VQFINAPR-DDRVVYNFHCYEDPLKFTHQGATWPTDIIIGERIMKFEDETSEAFVEELFST 242			
QY	238 IEEKSKKKRRIYIGEGAVRKADLESIKTFSVVERMEK-----RWSMAY 285			
DB	243 AISTPAKGTLLYCGEGLVDIVVAEDSLKAFKINNVFSKHGTSRCGKITRKWTSAF 300			

Search completed: April 22, 2003, 13:58:17
Job time : 16 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 13:55:54 ; Search time 85 seconds
(without alignments)
768,435 Million cell updates/sec

Title: US-09-880-729A-2

Perfect score: 1732
Sequence: 1 MGVDPFERNRKILGRGINIGN.....LRKTNKDLLEALIGDSIE 317

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_21:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertibrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1728	99.8	317	16 09X273	09X273 thermotoga
2	377	21.8	370	16 097KU0	097KU0 clostridium
3	367.5	21.2	414	2 059733	059733 rumiinosoccu
4	366	21.1	377	16 097KU1	097KU1 clostridium
5	363.5	21.0	637	2 059440	059440 fibrobacter
6	350.5	20.2	657	2 059446	059446 fibrobacter
7	330.5	19.1	1012	16 097DK6	097DK6 clostridium
8	328.5	19.0	332	2 060054	060054 unidentified
9	327.5	18.9	549	3 09Y870	09Y870 piromyces r
10	323	18.6	343	3 09AL36	09AL36 clostridium
11	319.5	18.4	1714	3 09P867	09P867 piromyces e
12	316.5	18.3	611	2 09LCC5	09LCC5 rumiinosoccu
13	315.5	18.2	669	2 059445	059445 fibrobacter
14	315	18.2	1232	3 059943	059943 neocallimast
15	312	18.0	477	3 09HFE0	09HFE0 orpiniomyces
16	311.5	18.0	477	3 013333	013333 orpiniomyces

17	308	17.8	362	3 09URH5	09URH5 piromyces r
18	305	17.6	482	3 001409	001409 neocallimast
19	302.5	17.5	388	3 013334	013334 orpiniomyces
20	298.5	17.2	329	16 09X274	09X274 thermotoga
21	297.5	17.2	574	16 09KF82	09KF82 bacillus ba
22	296.5	17.1	471	3 P78719	P78719 orpiniomyces
23	293	16.9	753	2 09LCC4	09LCC4 rumiinosoccu
24	283	16.3	400	2 008342	008342 bacillus sp
25	278	16.1	759	2 005143	005143 rumiinosoccu
26	277.5	16.0	356	5 09XXV3	09XXV3 epidiolum c
27	269.5	15.6	455	2 053302	053302 rumiinosoccu
28	268.5	15.5	1030	2 09XU99	09XU99 clostridium
29	264.5	15.3	322	2 09ZM63	09ZM63 prevotella
30	259.5	15.0	357	2 059441	059441 fibrobacter
31	259	15.0	519	2 047916	047916 fibrobacter
32	248.5	14.3	385	16 098F78	098F78 rhizobium 1
33	226.5	13.1	353	2 044878	044878 prevotella
34	226.5	13.1	924	2 006842	006842 prevotella
35	221	12.8	1449	10 08RU51	08RU51 oryza sativ
36	215.5	12.4	584	2 045397	045397 prevotella
37	205.5	11.9	397	2 08RP23	08RP23 paenibacill
38	195	11.3	660	2 09L3J2	09L3J2 clostridium
39	190	11.0	354	2 09F0G8	09F0G8 rhizobium m
40	182	10.5	452	16 097KK6	097KK6 clostridium
41	181.5	10.5	432	2 093R81	093R81 bacillus sp
42	180	10.4	727	2 09AF65	09AF65 clavibacter
43	179.5	10.4	424	16 098J61	098J61 rhizobium 1
44	175	10.1	389	2 059232	059232 bacillus sp
45	173	10.0	368	3 012637	012637 macrophomina

ALIGNMENTS

RESULT 1

ID 09X273 PRELIMINARY; PRT; 317 AA.

AC 09X273;

DT 01-NOV-1999 (Tremblrel. 12, Created)

DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)

DE ENDOGLUCANASE.

GN TM1751.

OS Thermotoga maritima.

OC Bacteria: Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.

OX NCBI_TaxID=2336.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MSB / DSM 3109;

RX MEDLINE=99287316; PubMed=10360571;

RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Utterback T.R., Malek J.A., Liner K.D., Garrett M.M., RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;

RT "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima";

RL Nature 399:323-329(1999).

DR EMBL; AEO01813; AAD36816.1; -;

DR HSSP; P07985; ICEC.

DR TIGR; TM1751; -;

DR InterPro; IPR001547; GH 5.

DR Pfam; PF00150; cellulase; 1.

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.

KW Complete proteome.

SQ SEQUENCE 317 AA; 37383 MW; F7DF20958A9A1435 CRC64;

Query Match 99.8%; Score 1728; DB 16; Length 317;

Best Local Similarity 99.7%; Pred. No. 3,1e-125;

Matches 316; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MGVDPFERNRKILGRGINIGNALEAPNEGDMGVYIKDEFIDIKRAGFSHVRIPIRWSTHA 60

```

Db      1 MGVDPEFNKRLTGRGINIGNALNEANEGDVGVIKDEFFDITKEAGFSHVRIPIRMSTHA 60
QY      61 YAPPPYKIMDRPEFKRVDEVYNGALKRGLAVAINIHHEELMNDPEEHKREFLAKQIAD 120
Db      61 YAPPPYKIMDRPEFKRVDEVYNGALKRGLAVAINIHHEELMNDPEEHKREFLAKQIAD 120
QY      121 RYKQVPELTFEELNEPIGNLTPEKKNELLEBAKVIRSIDKKNIIIGTAMGQISALE 180
Db      121 RYKQVPELTFEELNEPIGNLTPEKKNELLEBAKVIRSIDKKNIIIGTAMGQISALE 180
QY      181 KLSVPEKWEKNSIVTTHYNNPEFTTHOGAEWVGSEKWLGRKNGSPDDOKHLEEFNFTIE 240
Db      181 KLSVPEKWEKNSIVTTHYNNPEFTTHOGAEWVGSEKWLGRKNGSPDDOKHLEEFNFTIE 240
QY      241 WSKKKRPIYIGFEGAYRKADLESRIKWTSPYVREMERBRMSAWYEFCSGFGVYDTLTK 300
Db      241 WSKKKRPIYIGFEGAYRKADLESRIKWTSPYVREMERBRMSAWYEFCSGFGVYDTLTK 300
QY      301 TWNKDLLEALLIGDSIE 317
Db      301 TWNKDLLEALLIGDSIE 317

```

RESULT 2

```

Q97KU0      PRELIMINARY:      PRT:      370 AA.
AC      097KU0:
DT      01-OCT-2001 (Tremblrel. 18, Created)
DT      01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT      01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE      Endoglucanase family 5.
CN      CAC0826.
OS      Clostridium acetobutylicum.
OC      Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridia;
OC      Clostridiales; Clostridiaceae; Clostridium.
OX      NCBI_TaxID=1488;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-A7CC 824 / DSM 792 / VKM B-1787;
RX      MEDLINE=21359325; PubMed=1146286;
RA      Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA      Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,
RA      Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA      Bennett G.N., Koonin E.V., Smith D.R.;
RT      "Genome sequence and comparative analysis of the solvent-producing
RT      bacterium Clostridium acetobutylicum."
RL      J. Bacteriol. 183:4823-4838(2001).
DR      EMBL: AE007598; AAK78802.1; -.
DR      InterPro: IPR001547; GH_5.
DR      Pfam: PF00150; cellulase; 1.
DR      PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW      Complete proteome.
SQ      SEQUENCE 370 AA: 40934 MW: D47BDD23D2D6B92C CRC64;

```

Query Match

Best local Similarity .32.2%; Score 377; DB 16; Length 370;
Matches 102; Conservative 47; Mismatches 128; Indels 40; Gaps 10;

```

QY      15 GINIGNALA- -PNEGDMG- -VIKDEFFDITKEAGFSHVRIPIRMSTHAYAPPYKIMDRF 72
Db      51 GWNIGNTLIDASPDDETGMGNPKTYRAMIDRIKAGFNTRYIPVSWSHGAGSSTIIDQAW 110
QY      73 FKRVDEVYNGALKRGLAVAINIHHEELMNDPEEHKREFLAKQIADRYKDYPET 128
Db      111 LNRVOEVYNYIYQDMYALINTHTSWIIPYKNEASTDELTKVWGQIANRFXDYDSH 170
QY      129 LFFELNPPHGNLTPEKKNELLE- - - - -ALKVINSIDKKNH- - - - -IIGT- - - - -AEM 173
Db      171 LIFETLNPPRIYVSPEDWNGTAESRDVINKNFLTAVNTIRISTGNSNSRFLMPTVYAS 230
QY      174 GGISALELTVAPKWEKNSIVTTHYNNPEFTTHOGAEWVGSEKWLGRKNGSPDDOKHLE 233

```

```

Db      231 TATAAMDVLVPPNDRKVIYSLHWYAPYSF- - - - -AMPKGT- - - - -HWGCEADKDALDG 281
QY      234 EFNTEESKKNRPITYIGFEGAYRKADLESRIKWTSPYVREMERBRMSAWYEFCSG- - 291
Db      282 OLNAIYKRFYNGQPPVYIGEGSINKNESSRASLAKFYSDARKKGIITVWMD- -NGKS 339
QY      292 - - - - -FGVYDTLTKTW 302
Db      340 AVGDNDNGIILDRNNLTW 356

```

RESULT 3

```

Q59733      PRELIMINARY:      PRT:      414 AA.
AC      059733:
DT      01-NOV-1996 (Tremblrel. 01, Created)
DT      01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT      01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE      Beta-1,4-D-glucanase (EC 3.2.1.4).
CN      CELA.
OS      Ruminococcus albus.
OC      Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridia;
OC      Clostridiales; Lachnospiraceae; Ruminococcus.
OX      NCBI_TaxID=1264;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-Ar67;
RA      Vercoe P.E., Gregg K.;
RT      "Sequence and transcriptional analysis of an endoglucanase gene from
RT      Ruminococcus albus Ar67."
RL      Submitted (Feb-1993) to the EMBL/GenBank/DBJ databases.
DR      EMBL: L10243; AAA26467.1; -.
DR      HSP: P17901; IEDG.
DR      InterPro: IPR001547; GH_5.
DR      Pfam: PF00150; cellulase; 1.
DR      PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW      Glycosidase; Hydrolase.
SQ      SEQUENCE 414 AA: 45879 MW: F088CF0C41E3272B CRC64;

```

Query Match

Best local Similarity 30.7%; Score 367.5; DB 2; Length 414;
Matches 100; Conservative 51; Mismatches 132; Indels 43; Gaps 11;

```

QY      15 GINIGNALA- - - - -PNEGDMG- -VIKDEFFDITKEAGFSHVRIPIRMSTHAYAPPYK 67
Db      86 GWNIGSLATAGAGNASEVNMGNPKTYRKEMIDAYNNKGFYVIRIPYVWGSHVGGGPYK 145
QY      68 IMDRPEKRVDEVYNGALKRGLAVAINIHHEELMNDPEEH- - - - -KERFLAKQIADRYK 123
Db      146 IDENMLARVOEVYNYAVDQAVYIINSHEELRIPDNEHIDAVDEKTAIIMKOVAERFK 205
QY      124 DYETTFEELNPPHGNLTPEKKNELLEBAKVIRSI- - - - -DKKHTIIG 169
Db      206 DYGDHLFEELNPPHGNLTPEKKNELLEBAKVIRSI- - - - -DKKHTIIG 169
QY      170 T-AEWGQISALELTVAPKWEKNSIVTTHYNNPEFTTHOGAEWVGSEKWLGRKNGSPDDQ 228
Db      266 TVASSCGKLAIQDTAIP- - - - -DHIIGESIHAYTPYAFY- - - - -NANADWELFEMDNTPAT 318
QY      229 KHLIEENFTEESKKNRPITYIGFEGAYRKADL- - - - -ESRIKWTSPYVREMERK- - - - -WS 282
Db      319 AELITLMSNKENVLDIPDIPVITTEYGAVCKNNNEDEDRAKWSAVTLEYAEMLGIPCV- 377
QY      283 WAWYEFCSG- - - - -FGVYDTLTKTWKND 305
Db      378 WDNGYSSGNEELFGLFDRNCTWFTD 403

```

RESULT 4

```

Q97KU1      PRELIMINARY:      PRT:      377 AA.
AC      097KU1:
DT      01-OCT-2001 (Tremblrel. 18, Created)

```


Seq	SEQUENCE	657 AA:	73554 MW:	AA29F4A0FEB647 C8C64:
Q0	SEQUENCE	657 AA:	73554 MW:	AA29F4A0FEB647 C8C64:
Q1	Query Match	20.2%:	Score 350.5:	DB 2: Length 657:
Q2	Best Local Similarity	27.2%:	Pred. No. 9.8e-19:	
Q3	Matches	95: Conservative	57: Mismatches	142: Indels 55: Gaps 12:
Q4	10 KILGGINGNMLNAPNEDMGVIVKDEFFDIIEKAGFSHVRLPT---	PMSTHVAFAF---	63	
Q5	308 KTLNGLVNTNLENAADGFKSFELGEKAVOYLLAENGFSLSRLPDLIDLATNRDAFVWG	367		
Q6	64 --PPKYI--MDREFKRYDEVINGALKRGIAVAINIHHYELMNDPEEHKEFRLA----	WK 116		
Q7	368 TDTLKFDDDDTLFLVLDSEVWTKGNMSLVIDIHEYDMSYNTSVTKDNTNYIKMAEYWK	427		
Q8	117 QIADRYKDYR-ETLPEELINEP----	NGINLTPEKMNLLLEAKLVIRSIDKRRITITGAE 172		
Q9	428 HVAANHAESPREDLPEELINEPDMSDGKYTAOMVYAQAQMDAIDATVDTKHIIILGDQA	487		
Q10	173 MGSISALELSPVKKWKSIVTIHYVNPFEETHOGAEWE-----	GSEKW--- 217		
Q11	488 WYSISLLAK-RPFTODNIIVYIHTPEAFTHOGGSMWDTATIHDIPEPYDAKSTVS	546		
Q12	218 --LG-----	RMGSPDDCKHILEENFTLEEMSKKKRPYIEFGAYR- KAD 261		
Q13	547 GDFGVKSKRAYKVIKNIKNYTKGSEALIEQLLAKKMAATNNVPIITIEFGALNRST	606		
Q14	262 LESRIKTSFVYEMEKRRWMAVWEPCSGCYVDTLRKTWNKDLLEAL	310		
Q15	607 AESRIWVLAEMREICDTQLIPMTHMGVYTGNSFYIE-----	NGKLTGL 649		
Q16	RESULT 7			
Q17	Q97DK6	PRELIMINARY:	PRT: 1012 AA.	
Q18	AC	Q97DK6:		
Q19	DT	01-OCT-2001 (TREMBLrel. 18, Created)		
Q20	DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)		
Q21	DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
Q22	DE	Endoglycanase family 5, S-layer homology, cell-adhesion and dockerlin domains.		
Q23	GN	CAC3469.		
Q24	OS	Clostridium acetobutylicum.		
Q25	OC	Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridia;		
Q26	OX	Clostridiales; Clostridiaceae; Clostridium.		
Q27	NOBL_TaxID=1488:			
Q28	[1]			
Q29	RP	SEQUENCE FROM N.A.		
Q30	RC	STRAIN-ATCC 824 / DSM 792 / VKM B-11787:		
Q31	RX	MEDLINE=J1359325; Pubmed=11466286;		
Q32	RA	Neolling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,		
Q33	RA	Tatousov R.L., Sebache F., Doucet-Stein L., Soucaille P., Daly M.J.,		
Q34	RA	Benne G.N., Koonin E.V., Smith D.R.:		
Q35	RT	"Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum.";		
Q36	RL	J. Bacteriol. 183:4823-4838(2001).		
Q37	EMBL: AE007844: AAK81397.1:			
Q38	DR	InterPro: IPR003343: Big_2.		
Q39	DR	InterPro: IPR002105: Dockerin_1.		
Q40	DR	InterPro: IPR002048: EF-hand.		
Q41	DR	InterPro: IPR001547: GH_5.		
Q42	DR	InterPro: IPR001950: TIF_SUI1.		
Q43	DR	Pfam: PF02368: Big_2; 1.		
Q44	DR	Pfam: PF00404: Dockerin_1; 2.		
Q45	DR	PROSITE: PS00448: CLOS_CELULOSOME_RPT; UNKNOWN_1.		
Q46	DR	PROSITE: PS00018: EF_HAND; UNKNOWN_2.		
Q47	DR	PROSITE: PS00659: GLYCOSYL_HYDROL_F5; UNKNOWN_1.		
Q48	DR	PROSITE: PS01118: SUI1_1; UNKNOWN_1.		
Q49	KW	Complete proteome.		
Q50	SEQ	SEQUENCE 1012 AA: 110097 MW: A74C99ECC6FC9E45 C8C64:		

```

QY      2 GVDPERKKILGIRGINIGALE-APNEGDMG-VVIKDEFDIIKEGFSHVRPIRKSTH 59
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      383 GLSSMDILVKOKKQVMDGLGNSLESAGGETGWMGNPATRTKMKIDOKKAGFNVRIPRWDEH 442
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      60 AYAPRPYKIMRFRKRVDEVINGALKRGLAIVAINHHYEELMNPDEHKERFL-ALW 115
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      443 -YTSNVTIDIPAYRSRLETVVNYALANGMAIINHHINKLOGEAMNNAKMSVINEGSAIW 501
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      116 KOIADRYKDYPETLFEFFELINEPHGNTLPEKV-----NELLEALKYINSI---DK 162
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      502 KOVGHEFEDYDQGLPFTINEPR---TGDDMTGNSSSYNVNNEYNAKTLSTIRATGNGNG 558
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      163 KHTTITIG-----AEMGGISALEKLSVPKWEKNSIVTIHYNPFPTHOGAEWVGSEKWL 218
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      559 KRLVMPFYCASADAKAAVA--WVNP-DRVAVSHATIPYNA-MNDPTKGSISTF 613
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      219 GRKWSPPDOKHLEE-FNFIEMWSKKNRPRTYIGEGAYRKADLESRIKTSFVNVEME 277
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      614 G-----DADKACIDKTPRLDLKPTFKIGIYVVGEEFATKKNLODVPNAKYLLQVAS 667
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      278 KRKMSAWME-----FCSGGRVDTLRKTMN---KDLLEALIG 313
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      668 SYGIPCCWMDNNAFSPNSTDMSGIFN-RKTIQFYFPELVQAMLDG 711
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
060054 PRELIMINARY: PRT: 332 AA.
ID 060054
AC 060054
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Cellulase (EC 3.2.1.4).
DE CEL.
OS unidentified bacterium.
OC Bacteria; environmental samples.
OX NCBI_TaxID=2338;
RN RP SEQUENCE FROM N.A.
RX MEDLINE=96045463; PubMed=7546604;
RA Healy F.G., Ray R.M., Aldrich H.C., Wilkie A.C., Ingram L.O.,
RA Shannugam K.T.;
RT "Direct isolation of functional genes encoding cellulases from the
RT microbial consortia in a thermophilic, anaerobic digester maintained
RT on lignocellulose.";
RL Appl. Microbiol. Biotechnol. 43:667-674(1995).
DR EMBL: U12011; AAA91966.1; -.
DR HSSP: P07985; ICEN.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF00150; cellulase; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 332 AA; 38429 MW; 3D34FA90C39129C1 CRC64;

*Query Match 19.0%; Score 328.5; DB 2; Length 332;
Best Local Similarity 27.1%; Pred. No. 26-17;
Matches 92; Conservative 52; Mismatches 135; Indels 61; Gaps 11.

QY      14 RGINIGNAL-EAANE-GDMGVATKDEFPIKAGFSHVRPIRWSTHAYAPRPYK- 68
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      9 RGINIGWISQCPREOSHDSFTIGBEDIKTIAWGMGDHITLPT-----DYDVSD 58
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      69 -----MDFFRRVDEVINGALKRGLAIVAINTH-----HYEELMNPDEHKERFLAKM 116
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      59 DECKLTGIGIYVNDQVCMCRNNGNLITLDHKAQYAFYUNEDNTLTFESRVDQAFVNLK 118
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      117 QIADRYKDYPETLFEFFELINEPHGNTLPEKKNELLEELKTVRSIDKKHTTITIGAEWGI 176
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      119 EFARARYKYSDFVSEFELINEV-VDADGLMNDLIRRAITELRTAPRTKTVYVGIKRNKA 177
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      177 SALEKLSVPKWEKNSIVTIHYNPFPTHOGAEWVGSEKWLGRKWSPPDOKHLLIEFN 236
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

      ||| | : : : | : : | ||| | : : |
Db 178 STELLD-PEYDENIIFTHFYDPFLTHQGAWVKLPDVSLEYPSMDYRIAKELN 236
QY 237 FIE-----EWSKKNRPIYIGEGAVRKADLESRIKW---TSF 271
Db 237 QYDSQSAVAEMGRDYLVRMEPAIKTAEKHNVPYVYCEGVILISAPAECAVEYRDMHS 296
QY 272 VVREMEKRRMSAWMEFGSGGVYDTLRKTWNKDLEALI 311
Db 297 IFEEYGIGRAAWTYKSM--SEGVIN-----WPADEPQOII 329

```

RESULT 9

```

QY870 PRELIMINARY; PRT; 549 AA.
ID QY870;
AC QY870;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE Endoglucanase.
GN EGL-1.
OS Pirmomyces rhizinflatus.
OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
OC Neocallimastlicaceae; Pirmomyces.
OX NCBI_TaxID=73428;
RN RP SEQUENCE FROM N.A.
RC STRAIN=2301;
RA Liu J.-H., Prensost K., Qiu X., Selinger L.B., Cheng K.-I.,
RA Kawchuk L.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF165266; AAD43818.1; -
DR HSSP; P17901; IEDG.
DR InterPro; IPR002883; CBD_5.
DR InterPro; IPR001547; GH_5.
DR Pfam; PF02013; CBM_10; 3.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
SQ SEQUENCE 549 AA; 62631 MW; F2BC4A2530E7EF06 CRC64;

```

Query Match 18.9%; Score 327.5; DB 3; Length 549;

Best Local Similarity 29.1%; Pred. No. 4.6e-17; Matches 101; Conservative 50; Mismatches 139; Indels 57; Gaps 11;

```

QY 15 GINIGALEAP-----NEGDMG-VVTKDEFDIIKEAGSHVIRIPRMSTHA 60
Db 119 GMLGNLTDAECTSMNMYKKDPIGSECMGNVNTNEIYTLIDNCPNVRIPPTWTGHF 178
QY 61 YAFPPYKIMDRFRRVDEVINGALKRGLAVAINIHVEELMNDP---EHRERFLAW 115
Db 179 GEAPDYKINQMMKRVHEIYDYPKNGAFVILNIH--ETWNAHFTETLEEGVILIEKI 236
QY 116 KOIADRYKDYPTLFPFLINPEHGNLTPEKW-----NELLEALKVIRSID--- 161
Db 237 SQIAEEKDYDEHLIEGONEPRKNDTPVEMNGDGDGMVNVEMNAVFLKTIIRSSGNN 296
QY 162 -KKHTIIIGAEWGISALSKLSVPRKENSIVTIIHYNPEEFT-HOGAIVMEGSEKWL 219
Db 297 IKRHLMPRYAAAVNESLKNFNPDKDDKVIASGHAYSPYNALNNGAAVEYFDR--- 333
QY 220 RKGWSPDDQHLIEEFNIEEWSKKNRPIYIGEGAVRKADLESRIKWTSPVREMEKR 279
Db 354 -----SGMSELDMNVNLIKTFIDKIDPIVMGEYGMANRENEBERAKMAVEYMEKVTAL 407
QY 280 RMSAWMEFGSG-----EGYVDTLR-KTWKDLLEALIGDSTIE 317
Db 408 GIPOVWMD--NGVFESEGERFGLDRKNKIYVPSIVAALOKGKLE 452

```

RESULT 10

```

QY870 PRELIMINARY; PRT; 343 AA.
ID QY870;
AC QY870;

```

```

DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
DE Endo-1,4-glucanase (EC 3.2.1.4).
GN CELC.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1515;
RN RN SEQUENCE FROM N.A.
RC STRAIN=E7;
RA Zverlov V.V., Schwarz W.H.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ307315; CAC27410.1; -
DR HSSP; P07985; ICEN.
DR InterPro; IPR000463; Fatty_acid_BP.
DR InterPro; IPR001547; GH_5.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00214; FAPB; UNKNOWN_1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 343 AA; 40940 MW; BC7533D00A8268DE CRC64;

```

Query Match 18.6%; Score 323; DB 2; Length 343;

Best Local Similarity 28.1%; Pred. No. 5.5e-17; Matches 97; Conservative 61; Mismatches 125; Indels 62; Gaps 13;

```

QY 15 GINIG---MLEAPNEGDMGVVTKDEFDIIKEAGSHVIRIPRMSTHAYAFPPYKIMDR 71
Db 7 GINIGMISQYOVFSKHFPTITEKDIETIAEGFPHVRP-----FDYPIESDDN 59
QY 72 F-----FRVDEVINGALKRGLAVAINIH-----YELMNDPEHERFLAL 114
Db 60 VGEYKEDGLSYIDRCILMCKKYNGLVDIMHAGYRQDFKSTLEDPWQOK-REVDI 118
QY 115 WKQIADRYKDYPTLFPFLINPEHGNLTPE--KNELLEALKVIRSIDKKHTIIIGTAE 172
Db 119 WRFLAKRYINERHIAPELLNEV---VEPDSTRNKLIMLEYIKAIREIDSTMWLYIGNN 175
QY 173 WGSIALEKISVPRKENSIV-TIIHYNPEEFTHQGAENVBG-----SEKWLGRKGGSP 225
Db 176 YNSPDDLKILA--DIDDYIVYNFHFYNPFTHQKAHWSASAMAYNRYKVPQOYGIE 233
QY 226 DDOKH-----LIEEFNIE-----EWSKKNRPIYIGEGAVRKADLESRI 266
Db 234 EFKYKNNKYSFPMELNMLKLNKELRKDLKPAIEFRKKKKCKLTCGEFGYAIADLESRI 293
QY 267 KWTSEVVRERMEKRRMSAWMEFGS-GRGVYDTLRKTWNKDLEAL 310
Db 294 KMHEDYISLLEEDIGAVNNYKMKMDEIYNEDRKPVSQELVNL 338

```

RESULT 11

```

QY867 PRELIMINARY; PRT; 1714 AA.
ID QY867;
AC QY867;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE Endoglucanase 5A.
GN CEL5A.
OS Pirmomyces equi.
OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
OC Neocallimastlicaceae; Pirmomyces.
OX NCBI_TaxID=99929;
RN RN SEQUENCE FROM N.A.

```

```

RP MEDLINE=20391845; PubMed=10931904;
RA Eberhardt R.Y., Gilbert H.J., Hazlewood G.P.;
RT "Primary sequence and enzymatic properties of two modular
RT endoglucanases, Cel5A and Cel45A, from the anaerobic fungus Pirmomyces
RT equi.";

```


Db 483 KKHITLFGDAQWISISLIAK-RPPTDNDNIYVHTYEPFAFTHQGGSWTDYATIKGIPF 541
QY 213 --GSEKW-----LG-----RKWSPDDOKHLEEFNFIEEMSKKRPYI 251
Db 542 PYDAKSTYSGDGVKSTQSYKSNIKNYKSGKEALINVELLKKAKNAATNNVEYII 601
QY 252 GEFGAYR-KADLESRIKWTSEFVREMEKRRMSWAYWECGSGFYVD 296
Db 602 NEFGALMLRSTADRLNLYFTAMREICDTLQIPWTHWGYTGWFSLE 647

RESULT 14

059943 PRELIMINARY; PRT; 1232 AA.
AC 059943;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Cellulase Celld (Fragment).
GN CELD.
OS Neocallimastix patriciarum (Rumen fungus).
OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
OC Neocallimasticeae; Neocallimastix.
OX NCBI_TaxID=4758;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93123992; Pubmed=1479358;
RT "A novel polysaccharide hydrolase cDNA (celld) from Neocallimastix patriciarum encoding three multi-functional catalytic domains with high endoglucanase, cellobiohydrolase and xylanase activities.";
RL J. Gen. Microbiol. 138:0-0(0).
DR EMBL/AF053363; AAC06321.1;
DR HSSP: P17901; 1EDG.
DR InterPro: IPR002883; CBD_5.
DR InterPro: IPR001547; GH_5.
DR Pfam: PF02013; CBM_10; 2.
DR Pfam: PF00150; cellulase; 3.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_3.
FT NON TER
SQ SEQUENCE 1232 AA; 140617 MW; E58133110F684348 CRC64;

Query Match 18.2%; Score 315; DB 3; Length 1232;
Best Local Similarity 27.4%; Pred. No. 1.2e-15;
Matches 87; Conservative 56; Mismatches 120; Indels 54; Gaps 10;

QY 10 KILGRGINIGNALEA-----PNEGDMG-VVINDEFDIIKEAGFSVIRIPR 55
Db 798 KEMFGNNLNTLDAOCIEYLNKDQDTASETCWGNPKTTEDEKVLMDNQFNFRIPPT 857
QY 56 WSTHAYAFPPYKIMDRFFKRVDEYINGALRGGLAVAINIH-----YEELMNDPEEHE 109
Db 858 WSGHFGAPRYKINEKLVKHEIVDPRYKNGAFVILNLHETMNAFSETLDIAKILE 917
QY 110 RFLALMKQIDRYKDYETLEFELINPHGNLPPEKVN-----NELLEALKVIR 158
Db 918 K---IMSQIAEEFDYDEHILFEGLINPRKNDPVEWTGDDQEGDAVNMNNAVFLKTIR 974
QY 159 SI-----DKKTIIGTLEMWGISALEKLSVPMKKNSTIVIIHYNPEFT-HOGAEVVEG 213
Db 975 SSGGNPKKRLMLPRYAAACNENSKNFPEDDDKVIASVHAAPYFALNNGAGAVD 1034
QY 214 SEKWIGRK--WCSPPDOKHLEEFNFIEEMSKKRPYIGFAGAYRKADLESRIKWT 270
Db 1035 FDA-AGKKDELMNINLMKKRFVDQ-----GIPMLIGEGAMNRRNEERATWAE 1082
QY 271 FVYREMEKRRMSWAYW 287
Db 1083 FYMEKVTAMGVPOVWMD 1099

RESULT 15
Q9HFE0

ID Q9HFE0 PRELIMINARY; PRT; 477 AA.
AC Q9HFE0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE 1,4-Delta-D-glucan-4-glucanohydrolase (EC 3.2.1.4).
GN CELD.
OS Orphiomyces sp. PC-2.
OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;
OC Neocallimasticeae; Orphiomyces.
OX NCBI_TaxID=50059;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PC-2;
RX MEDLINE=98146543; Pubmed=9485595;
RA Chen H., Li X.L., Blum D.L., Ljungdahl L.G.;
RT "Two genes of the anaerobic fungus Orphiomyces sp. strain PC-2 encoding cellulases with endoglucanase activities may have arisen by gene duplication.";
RL FEMS Microbiol. Lett. 159:63-68(1998).
DR EMBL: U97153; AAC05164.1; .
DR HSSP: P17901; 1EDG.
DR InterPro: IPR002883; CBD_5.
DR InterPro: IPR001547; GH_5.
DR InterPro: IPR001230; Frenyl_site.
DR Pfam: PF02013; CBM_10; 2.
DR Pfam: PF00150; cellulase; 1.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.
DR PROSITE: PS00294; PRENITATION; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 477 AA; 53635 MW; 6CBEC85A469BA0DA CRC64;

Query Match 18.0%; Score 312; DB 3; Length 477;
Best Local Similarity 27.9%; Pred. No. 5.9e-16;
Matches 96; Conservative 49; Mismatches 115; Indels 84; Gaps 14;

QY 10 KILGRGINIGNALRAP--NEGDMGVNIRKDE-----FPDIKEAGFSHVRI 52
Db 31 KEINIGNNLGNALDAHCLDKLDYN--KDQLASETCWGNPKATPGLSALKNGFNFR 87
QY 53 PIRKSTHAYAFPPYKIMDRFFKRVDEYINGALRGGLAVAINIH-----ELMNDPEEHE 109
Db 88 PTTWTGHHGNGPDKYKIDVMMRRVHEVVDYALNTGSYVILNIHNNMNYAFSNNLOKAKR 147
QY 110 RFLALMKQIDRYKDYETLEFELINPHGNLPPEKKNELLEALKVIRSDKKHTIIG 169
Db 148 ILAALIMKQIAAEFANYDEHILFEGLINPRKVDHPNENNGDQEGWDF--VNEKNVFLQ 204
QY 170 TAE-WGISALEKLSVPMKKNSTIVIIHYNPEEF--THQ 206
Db 205 TVRASGNNMIRHLMPTIYACVNDALLESYVRKFTPTNDKVIASVSYVPYNALNTGA 264
QY 207 GAENVESSEKWLGRKWSPPDOKHLEEFNFIEEMSKK-KR-----PIYICEFGAYR 258
Db 265 GAE-----KTFGSTSD-----IEMANNKIKRFLVDRNIPVIGIFEGAMN 303
QY 259 KADLESRIKWTSEFVREMEKRRMSWAYW--FCSG-----FGVD 296
Db 304 RDNESERARMAEYIKSATAMGVPCVILMDNGITOGTGELEFVID 347

Search completed: April 22, 2003, 13:59:51
Job time : 89 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 13:57:14 ; Search time 29 Seconds
(without alignments)
321.623 Million cell updates/sec

Title: US-09-880-729A-2

Sequence: 1 MGVDPFERNKILGRGINICN.....LRTKWNKDLLEALIGDSIE 317

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/FOCUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1732	100.0	317	2	US-09-066-075-2
2	1732	100.0	317	2	US-08-518-615A-2
3	1732	100.0	317	3	US-08-951-889-2
4	1732	100.0	317	4	US-09-472-857-2
5	835.5	48.2	360	4	US-09-134-078-27
6	312	18.0	477	3	US-09-118-324-2
7	311.5	18.0	473	4	US-09-286-691-15
8	311.5	18.0	473	4	US-09-687-147-15
9	308	17.8	394	4	US-09-419-459-4
10	308	17.8	390	4	US-09-419-459-10
11	308	17.8	582	4	US-09-419-459-2
12	306.5	17.7	574	3	US-08-732-433-1
13	306.5	17.7	574	4	US-08-945-574-2
14	296.5	17.1	471	3	US-09-118-324-5
15	296.5	17.1	471	4	US-09-286-691-12
16	296.5	17.1	471	4	US-09-687-147-12
17	281.5	16.3	537	4	US-09-110-959A-4
18	265.5	15.3	566	2	US-07-862-588B-4
19	177	10.2	467	2	US-08-727-548-2
20	177	10.2	467	4	US-08-945-574-1
21	170	9.8	352	4	US-09-102-204-1
22	168	9.7	400	2	US-08-713-298B-2
23	168	9.7	400	3	US-08-870-180B-2
24	168	9.7	400	3	US-08-814-052-4
25	168	9.7	400	3	US-08-812-829-4
26	168	9.7	400	4	US-09-226-529-2
27	168	9.7	462	2	US-08-870-180B-13

28	168	9.7	462	4	US-09-226-529-13	Sequence 13, Appl
29	159	9.2	551	2	US-09-033-537A-1	Sequence 1, Appl1
30	145.5	8.4	430	2	US-08-924-440-2	Sequence 2, Appl1
31	143.5	8.3	956	4	US-09-134-078-63	Sequence 63, Appl1
32	140	8.1	616	4	US-09-136-574A-47	Sequence 47, Appl
33	140	8.1	1426	4	US-09-136-574A-43	Sequence 43, Appl
34	139.5	8.1	327	2	US-08-169-948B-16	Sequence 16, Appl
35	139.5	8.1	327	2	US-08-448-873-16	Sequence 16, Appl
36	139.5	8.1	327	4	US-08-382-452D-16	Sequence 16, Appl
37	130.5	7.5	418	4	US-09-254-733-5	Sequence 5, Appl1
38	124	7.2	613	4	US-09-149-727-5	Sequence 5, Appl1
39	123.5	7.1	553	1	US-08-651-572-2	Sequence 2, Appl1
40	123.5	7.1	553	3	US-09-066-544-2	Sequence 2, Appl1
41	123.5	7.1	553	3	US-08-951-086-2	Sequence 2, Appl1
42	123.5	7.1	553	4	US-09-430-669-2	Sequence 2, Appl1
43	121	7.0	663	4	US-09-134-078-61	Sequence 61, Appl
44	121	7.0	680	4	US-09-134-078-25	Sequence 25, Appl
45	117.5	6.8	531	2	US-07-862-588B-7	Sequence 7, Appl1

ALIGNMENTS

```
RESULT 1
US-09-066-075-2
; Sequence 2, Application US/09066075
; Patent No. 5925749
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,075
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615
; FILING DATE: August 23, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-09-066-075-2

Query Match      100.0%  Score 1732:  DB 2:  Length 317:
Best Local Similarity 100.0%:  Pred No. 6, 2e-161:
Matches 317:  Conservative 0:  Mismatches 0:  Indels 0:  Gaps 0:
OY      1 MGVDPFERNKILGRGINIGALAPNCGDWGVYIKDFPIIKKAGFSHWIRIPIRNSTHA 60
|||||
```

Db 1 MGVDPEERNKILRGINIGNALPNEGDWGVVTKDEFFDIKEAGFSHVRIPIRWSTHA 60
QY YAFPPYKIMDRPFKRVDEVINGALKRGIAVAIINHHEELMNDPEHKEFFLALMKQIAD 120
Db 61 YAFPPYKIMDRPFKRVDEVINGALKRGIAVAIINHHEELMNDPEHKEFFLALMKQIAD 120
QY 121 RYKDYPELFEILNEPFGNLTPEKWNELLEALKVIRSIDKHTIIGTAEMGGSIALE 180
Db 121 RYKDYPELFEILNEPFGNLTPEKWNELLEALKVIRSIDKHTIIGTAEMGGSIALE 180
QY 181 KLSVPMKWNKNSIVTIHYNPFETHOGAEWVGSEKMLGRKWSGSPDDOKHLEEFNFTIE 240
Db 181 KLSVPMKWNKNSIVTIHYNPFETHOGAEWVGSEKMLGRKWSGSPDDOKHLEEFNFTIE 240
QY 241 WSKKRRPIYIGEGFAYRKADLESRIKWTSPVVRMEKRRMSAWYEFCSGFGVYDTLRK 300
Db 241 WSKKRRPIYIGEGFAYRKADLESRIKWTSPVVRMEKRRMSAWYEFCSGFGVYDTLRK 300
QY 301 TWNKDLLEALIGDSIE 317
Db 301 TWNKDLLEALIGDSIE 317

RESULT 2

US-08-518-615A-2
; Sequence 2, Application US/08518615A
; Patent No. 5962258
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615A
; FILING DATE: August 23, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ. ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-518-615A-2

Query Match 100.0%; Score 1732; DB 2; Length 317;
Best Local Similarity 100.0%; Pred. No. 6.2e-161;
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGVDPEERNKILRGINIGNALPNEGDWGVVTKDEFFDIKEAGFSHVRIPIRWSTHA 60
|||||

Db 1 MGVDPEERNKILRGINIGNALPNEGDWGVVTKDEFFDIKEAGFSHVRIPIRWSTHA 60
QY 61 YAFPPYKIMDRPFKRVDEVINGALKRGIAVAIINHHEELMNDPEHKEFFLALMKQIAD 120
Db 61 YAFPPYKIMDRPFKRVDEVINGALKRGIAVAIINHHEELMNDPEHKEFFLALMKQIAD 120
QY 121 RYKDYPELFEILNEPFGNLTPEKWNELLEALKVIRSIDKHTIIGTAEMGGSIALE 180
Db 121 RYKDYPELFEILNEPFGNLTPEKWNELLEALKVIRSIDKHTIIGTAEMGGSIALE 180
QY 181 KLSVPMKWNKNSIVTIHYNPFETHOGAEWVGSEKMLGRKWSGSPDDOKHLEEFNFTIE 240
Db 181 KLSVPMKWNKNSIVTIHYNPFETHOGAEWVGSEKMLGRKWSGSPDDOKHLEEFNFTIE 240
QY 241 WSKKRRPIYIGEGFAYRKADLESRIKWTSPVVRMEKRRMSAWYEFCSGFGVYDTLRK 300
Db 241 WSKKRRPIYIGEGFAYRKADLESRIKWTSPVVRMEKRRMSAWYEFCSGFGVYDTLRK 300
QY 301 TWNKDLLEALIGDSIE 317
Db 301 TWNKDLLEALIGDSIE 317

RESULT 3

US-08-951-889-2
; Sequence 2, Application US/08951889
; Patent No. 6008032
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,889
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/518,615
; FILING DATE: August 23, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ. ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-951-889-2

Query Match 100.0%; Score 1732; DB 3; Length 317;
Best Local Similarity 100.0%; Pred. No. 6.2e-161;
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGVDPEERNKILRGINIGNALPNEGDWGVVTKDEFFDIKEAGFSHVRIPIRWSTHA 60
|||||


```

|||||
Db 1 MGVDPEFRNKILRGINIGNALPNEGDMGVYIKDEFFDILKEAGSHRIPDIRMSTHA 60
OY 61 YAEPPYKIMDRFEKRVDEVINGALRKGLAVAINIHHEEELMNDPEEKERFLAMKOIAD 120
Db 61 YAEPPYKIMDRFEKRVDEVINGALRKGLAVAINIHHEEELMNDPEEKERFLAMKOIAD 120
OY 121 RYKDYPETLFEFELINEPHGNLTPPEKNMELLEALKVIRSIDKKHTIIGTAEMGGSISALE 180
Db 121 RYKDYPETLFEFELINEPHGNLTPPEKNMELLEALKVIRSIDKKHTIIGTAEMGGSISALE 180
OY 181 KLSVPKKEKNSIYTIHYNNPEFTHOGAEWVEGSEKWLGRKWSPPDQKHLIEEFNIEE 240
Db 181 KLSVPKKEKNSIYTIHYNNPEFTHOGAEWVEGSEKWLGRKWSPPDQKHLIEEFNIEE 240
OY 241 WSKKNRPYIIGFEGAYRKADLESRIKWTSPVYREMEKRRMSNAWBFCSGFGVYDTLRK 300
Db 241 WSKKNRPYIIGFEGAYRKADLESRIKWTSPVYREMEKRRMSNAWBFCSGFGVYDTLRK 300
OY 301 TWNKDLLEALIGDSTE 317
Db 301 TWNKDLLEALIGDSTE 317

```

```

RESULT 4
US-09-472-857-2
; Sequence 2, Application US/09472857
; Patent No. 6245547
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from
; TITLE OF INVENTION: Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/472,857
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/951,889
; FILING DATE:
; APPLICATION NUMBER: 08/518,615
; FILING DATE: August 23, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-09-472-857-2
Query Match 100.0%; Score 1732; DB 4; Length 317;
Best Local Similarity 100.0%; Pred. No. 6,2e-161;

```

```

Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MGVDPEFRNKILRGINIGNALPNEGDMGVYIKDEFFDILKEAGSHRIPDIRMSTHA 60
Db 1 MGVDPEFRNKILRGINIGNALPNEGDMGVYIKDEFFDILKEAGSHRIPDIRMSTHA 60
OY 61 YAEPPYKIMDRFEKRVDEVINGALRKGLAVAINIHHEEELMNDPEEKERFLAMKOIAD 120
Db 61 YAEPPYKIMDRFEKRVDEVINGALRKGLAVAINIHHEEELMNDPEEKERFLAMKOIAD 120
OY 121 RYKDYPETLFEFELINEPHGNLTPPEKNMELLEALKVIRSIDKKHTIIGTAEMGGSISALE 180
Db 121 RYKDYPETLFEFELINEPHGNLTPPEKNMELLEALKVIRSIDKKHTIIGTAEMGGSISALE 180
OY 181 KLSVPKKEKNSIYTIHYNNPEFTHOGAEWVEGSEKWLGRKWSPPDQKHLIEEFNIEE 240
Db 181 KLSVPKKEKNSIYTIHYNNPEFTHOGAEWVEGSEKWLGRKWSPPDQKHLIEEFNIEE 240
OY 241 WSKKNRPYIIGFEGAYRKADLESRIKWTSPVYREMEKRRMSNAWBFCSGFGVYDTLRK 300
Db 241 WSKKNRPYIIGFEGAYRKADLESRIKWTSPVYREMEKRRMSNAWBFCSGFGVYDTLRK 300
OY 301 TWNKDLLEALIGDSTE 317
Db 301 TWNKDLLEALIGDSTE 317

```

```

RESULT 5
US-09-134-078-27
; Sequence 27, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hallie, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1456
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-134-078-27
Query Match 48.2%; Score 835.5; DB 4; Length 360;

```

Best Local Similarity 49.5%; Pred. No. 2e-73;
Matches 154; Conservative 53; Mismatches 95; Indels 9; Gaps 3;

Qy	6	FERNKIIIGRGINIGNALNEANEDMGVJKDEPFDDIIKAGFESHVPIPRMSHNAAPP	65
Db	51	FEYMKMGKGVGNIGNALNEAPEGAMGVRIEDFEIILKKGDFSVAPILPRMSHISEKPP	110
Qy	66	YKIDREFKRRVDEVINGALRGLAVATINIHYEELMNDPEHKERFLALMKOJADRYKTY	125
Db	111	YDIDRNFEBERNAHVADALEENMLVLIINTHHPELTYOEPKVDYVETLWROIARFEEKY	170
Qy	126	PETLPELLEINPHGNLTPREPKMNETLEBALVYISIDKHTIIIGTAEWGISALEKLSVP	185
Db	171	PENLPEFELYNPDAONLTAERKNALPYKVLVLIRESNPTRIIIDAPMAHYSVRSKLTLY	230
Qy	186	KWEKNSIVTTHIYVNPDEETHOGAEWVGC-----SKWLGKRWGSPDDOKHILIEEFNIEEM	241
Db	231	N-DKRIIVSFHYIEFEKTHOGAAMVNPIPVRYKMGMEETINQJLSH---FKVYSDB	285
Qy	242	SKKNRPYIEIEEFGAYRKADLESRIKWTSPFYRREMERKRMAMWEEFCSGEYDTLIRKT	301
Db	286	AKQNNVPFLTELEFGAYSADMSVKTWTESRYKMAEEFGSYAVWEFCAGFGIYDRWSON	345
Qy	302	WNKDLLEALLIG 312	
Db	346	WIEPLATAVVG 356	

RESULT 6
US-09-118-324-2
Sequence 2, Application US/09118324

```

: Patent No 6110720
: GENERAL INFORMATION:
: APPLICANT: Li, Xin-Liang
: APPLICANT: Ljungdahl, Lars G.
: APPLICANT: Chen, Huizhong
: TITLE OF INVENTION: Orp1nomyces Cellulase CelE Protein and Coding Sequences
: FILE REFERENCE: 32-98sequence 1stling
: CURRENT APPLICATION NUMBER: US/09/118,324
: CURRENT FILING DATE: 1998-07-17
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: PatentIn Ver. 2.0
: SEQ. ID NO. 2
: LENGTH: 477
: TYPE: PRT
: ORGANISM: Orp1nomyces sp. PC-2
: US-09-118-324-2

```

```

Db      304 RDNESERARWAEYIKSATAMGVPCLWMDNGYTQGTGELFGVID 347
          : : | : | : : : | : : | ||| |

```

RESULT 7
 US-09-286-691-15
 Sequence 15, Application US/09286691
 Patent No. 6190189
 GENERAL INFORMATION:
 APPLICANT: Li, Xin-Liang
 APPLICANT: Li, Xunghong
 APPLICANT: Chen, Huizhong
 TITLE OF INVENTION: Cellulases and Coding Sequences
 FILE REFERENCE: 42-96
 CURRENT APPLICATION NUMBER: US/09/286,691
 CURRENT FILING DATE: 1999-04-05
 EARLIER APPLICATION NUMBER: US 60/027,883
 EARLIER FILING DATE: 1996-10-04
 EARLIER APPLICATION NUMBER: PCT US97/18008
 EARLIER FILING DATE: 1997-10-03
 NUMBER OF SEQ ID NOS: 29
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 15
 LENGTH: 473
 TYPE: PRT
 ORGANISM: Neocallimastix patriciarum
 US-09-286-691-15

	Query Match	18.0%;	Score 311.5;	DB: 4;	Length 473;	
	Best Local Similarity	28.0%;	Pred. No. 3.6e-22;			
	Matches 98; Conservative	53;	Mismatches 136;	Indels 63;	Gaps 13;	
Qy	10 KILGGINGNLBA-----PNEGDMG-VVIKDEFPIIIEKAFSHVRIPIR 55	:	:	:	:	:
Dd	30 KOLLITWISGNTLDTCETTEDLYDNKNQLASLEWCNMVATTDGLYKYKLSDLFENFRIPPT 89	:	:	:	:	:
Qy	56 WSTHVAAPBPKIMDFEERKDEVDINGALRGLAVAINIH---YEELMDPREHEKEFL 112	:	:	:	:	:
Dd	90 MEGHGFNADPYKINDQMRRKRVHEIYDVALINTGGYAALINHETWNHAFOKMLESAKTLV 149	:	:	:	:	:
Qy	113 ALMKQADADYKKOYPETLPFELLNEPHNGLTPEKM-----NELLEALKIRSID 161	:	:	:	:	:
Dd	150 AWMQJLAFAEDYDELTLTEGMNPERKAGDPRAEMNGDGEYNFPVENKNDLFVKIRRTG 209	:	:	:	:	:
Qy	162 ----KKHTIIIGTAEMGGSIALEKLSVPMEKNSIVTIIHYNPFEET-HOGA----EWE 212	:	:	:	:	:
Dd	210 GNNALRLHLMLPYAACINCINALINNRKFGDDKVILSLHSYSPYNDALENAGAISNFYD 269	:	:	:	:	:
Qy	213 GSE-KWLGRKWSPPDDQKHLIEEFNLIEEMSKSNKRPRITYIGFEAGVAKRADLESRIKTWF 271	:	:	:	:	:
Dd	270 GSEIDW-----AMNTINSKFISRGIPIVIGFEGVMANRNNDREHMAEY 313	:	:	:	:	:
Qy	272 VVAEMEKRRMSNAWE--PCSG---FGYYD--TLTKTNKKLLALLIG 313	:	:	:	:	:
Dd	314 YIKRASIGPCVIWDNGTFEEGEGEFGGLINSTLOAYTPK-LVNGCLIKG 362	:	:	:	:	:

Qy	53	PIKWSHAYAAFPYKYIMDFEKKRVDEYINGALKRGLAVAINTHNE---	ELMNDPEEIKE	109
Db	88	PTWTGHFGNGBPYKISIDWMMRVHEVVDYALNTSSYLINHHNNMVAFSNNLOKAP		147
Qy	110	RFALMKOTADRYKDYPELTFEELINEPBGNLTPKMYNLELEAKVRISIDKHTIIG		165
Db	148	ILAIIMQIDAEFANIDHEILIFEGMNEPKRVHPNEMNGDDEGDF---	VENEMNAVLO	204
Qy	170	TAE-WGISALEKESYPKW----	EKNSIVTHIYNPEE---	THO 204
Db	205	TVASGONGNNAIRLMLPTVAAACVNDGALSYARKFPTDNOKYIASHSVYPENFLNTGA		264
Qy	207	GAEMWEGSEWGLGRKNGSPDDCKHLLIEFNFIEMSKU-KR-----	PIYIGEGAYR	254
Db	265	GAE-----KTEGTSJD-----	IEMAMNKKFPLVDNRNIVLIGEGAMN	304
Qy	259	KADLESRIKTSFVVREREMERKRWMAWYE--FQSG----	FGYVD	296


```

; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Chang, Chia-Chieh
; TITLE OF INVENTION: CELLULASE ENZYMES
; FILE REFERENCE: 08919/037001
; CURRENT APPLICATION NUMBER: US/09/419,459
; CURRENT FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Piromyces rhizinflata
; US-09-419-459-2

```

```

Query Match          17.8%; Score 308; DB 4; Length 582;
Best Local Similarity 26.3%; Pred. No. 1, 1e-21;
Matches 83; Conservative 54; Mismatches 127; Indels 52; Gaps 8;

```

```

QY 10 KILGGINIGALEA-----PNEGDCG-VVTKDEFDIIKAGFSHVRIPIR 55
DB 148 KEMNFGMNLGNTMDACLEYLYNEKDOTASETCWGNPKTTEMFKVLIDNCFVFRIPPT 207
QY 56 WSTHVAAPRYKIMDFEKRVEVINGALKRGIAVAIINH-----YELMNDPEHKE 109
DB 208 WSGHGEADPYKIDKMKLRVHEVDYPRKNGAFVILNLHETMNAFSETIDTAKELLE 267
QY 110 RFLALMKQIADRYKDYPTLFFELINEPHCNLTPEKW-----NELLEALKVIR 158
DB 268 K---IMSQIAEEKDYDEHLIEGLNEPRKNDTPVEMTGDDCGMDAVANMNVFLKTVR 324
QY 159 SI---DKHTTIIIGTAEWNGISALEKLSVPMKKNISYTHIYNPEEFTHOGAEWVEGS 214
DB 325 SAGGNPKRHLMPYAAACNENSNFNFTPEDDDKVIASVHAAYNPNALNNGEGAVDK 384
QY 215 EKWLGRK---WSPDDQKHLIEEFNIEEMSKNKRPIYIGEGAYRKADLESRIKWTSF 271
DB 385 FPAAGRKDELMNINLMKKRRVDO-----GIPMLIGEGAMNRDNEEDRATWAER 433
QY 272 VVREMEKRRMSWAYME 287
DB 434 YMEKVTAMGVPOIWM 449

```

```

RESULT 12
; US-08-732-433-1
; Sequence 1, Application US/08732433
; Patent No. 6063611
; GENERAL INFORMATION:
; APPLICANT: van Solingen, Dieter
; TITLE OF INVENTION: No. 6063611el Alkaline Cellulase and
; TITLE OF INVENTION: Method of Producing Same
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/732,433
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05651
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L
; REGISTRATION NUMBER: 35,696

```

```

; REFERENCE/DOCKET NUMBER: GC287-US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-732-433-1

```

```

Query Match          17.7%; Score 306.5; DB 3; Length 574;
Best Local Similarity 26.0%; Pred. No. 1, 5e-21;
Matches 87; Conservative 57; Mismatches 126; Indels 65; Gaps 12;

```

```

QY 15 GINIGNALEA--PNEGDCG-VVTKDEFDIIKAGFSHVRIPIRWSHVAAPRYKIMOR 71
DB 45 GNNLGNTEFVAGQDETANGNPRVTRELIERIADGYSIRIPVTWENRIGGADYPIDPO 104
QY 72 FFKRYDEVINGALKRGIAVAIINHIELMNDPEHKE-----ERFLAMKQIADRYKDYR 126
DB 105 FLNRVDEVVQWALBEDLYVMTNLHDSWLVMTYEMHNNGVMKXRYSLMQLSHHEDYP 164
QY 127 ETLPFEELINEPHCNLTPEKMWELLE--EALKVIRSIDKHTIIG--TAENGISALEKLS 183
DB 165 TKLMFESVNEPKFS---QWGEIRRNHAL-----LDDLNTVFPEIIVROSOGONDYRPLV 216
QY 184 VKRWE-----KNSITVTHIYNPEEFTHOGAEWVEGSEKKLGKRWG 223
DB 217 LPTMETATSOPLNMLYOTIDKLDDPNLATVTHYIGFPEFSVNIAGITREE----- 268
QY 224 SPDDQKHLIEEFNIEEMSKNKRPIYIGERG--AYRK---ADLESRIKWTSVVEME 277
DB 269 --DSKREIETIDRYHNHTVANGIPVVLGEGGLGFDKHTGVTGQGEKLFKEFYLLHNL 326
QY 278 KRRMSWAYMEFCSGRG-----VYDTLAKTV 302
DB 327 ERDITHMLDNGQHFNRTYEWYDEELDMLRASW 361

```

```

RESULT 13
; US-08-945-574-2
; Sequence 2, Application US/08945574
; Patent No. 6313081
; GENERAL INFORMATION:
; APPLICANT: Ienting, Hermanus Bernardus Maria
; APPLICANT: Maurel, Karl-Heinz
; APPLICANT: Kottwitz, Beatrix
; APPLICANT: Weiss, Albrecht
; APPLICANT: van Solingen, Pieter
; TITLE OF INVENTION: Detergents Comprising Cellulases
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 140 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: MS Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/945,574
; FILING DATE: unavailable
; CLASSIFICATION:
; CLASSIFICATION: C11D 3/386
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP96/01755

```

FILING DATE: 26 Apr. 1996
 APPLICATION NUMBER: EP 95201115.3
 FILING DATE: 28 Apr. 1995
 APPLICATION NUMBER: U.S. 614,115
 FILING DATE: 12 Mar. 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy, Glenn E. J.
 REGISTRATION NUMBER: 33,539
 REFERENCE/DOCKET NUMBER: H 1920 PCT/US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (610) 832-2228
 TELEFAX: (610) 941-6067
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 574 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-945-574-2

Query Match 17.7% Score 306.5; DB 4; Length 574;
 Best Local Similarity 26.0%; Pred. No. 1.5e-21;
 Matches 87; Conservative 57; Mismatches 126; Indels 65; Gaps 12;

QY 15 GINIGNALAE--PNECDWG--VVIDEFDIIEAGFSHVIRIPIMSTHAYAFPPYKIMDR 71
 DB 45 GWNIGTFEDAVGOETAMGNPRVRELIERADGYSKIRIPVWERNIGAPYIPIDPQ 104
 QY 72 FFKVDEVINGALKRGALVAIINHIEELMNDPEEHK-----EFLALMKQIADRYDYP 126
 DB 105 FLNKDEVYQWALKEEDLYVMINLHDSMLYEMEHYNGVMAYKRSIMEDLSNHFDP 164
 QY 127 ETLFEELINEPHGNLTDEKNNELLE--EALKVIRSIDKKHIIIG--TAEMGSIALKLS 183
 DB 165 TKLMFESVNEPKES---QNMGEIRENNHAL-----LDDLNVFEFYIRQSGGQNDIRPLV 216
 QY 184 VPKWE-----KNSIVTIHYNPEEFTHOGAVEGSEKMLGRKWG 223
 DB 217 LPTWETATSQPLNNLYQITDKLDDPNLIATVHYHGFWSPSVNAGTTRFEE----- 268
 QY 224 SPDDQKHLIEEFNFIEEMSKKRRPIYIGERG--AYRK---ADLESRIKWTSVVAREME 277
 DB 269 --DSKRRIETFFDVHHTFVARGIPVVLGEGGLGFDKHTGVIOGGRKLFEEYLIIHLN 326
 QY 278 KRMSMAYWEFCGSG-----YVDPLRKTW 302
 DB 327 ERDITHMLMDNGOHFNHRTYEWIDEELFMDLRASW 361

RESULT 14
 US-09-118-324-5
 Sequence 5, Application US/09118324
 Patent No. 6110720
 GENERAL INFORMATION:
 APPLICANT: Li, Xin-Liang
 APPLICANT: Ljungdahl, Lars G.
 APPLICANT: Chen, Huizhong
 TITLE OF INVENTION: Orpinomyces Cellulase Celc Protein and Coding Sequences
 FILE REFERENCE: 32-98Sequence Listing
 CURRENT APPLICATION NUMBER: US/09/118,324
 NUMBER OF SEQ ID NOS: 5
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 5
 LENGTH: 471
 TYPE: PRT
 ORGANISM: Orpinomyces sp. PC-2
 US-09-118-324-5

Query Match 17.1% Score 296.5; DB 3; Length 471;
 Best Local Similarity 28.0%; Pred. No. 1e-20;
 Matches 98; Conservative 51; Mismatches 134; Indels 67; Gaps 14;

QY 10 KILRGINIGNALAE-----PNECDWG--VVIDEFDIIEAGFSHVIRIPIR 55
 DB 30 KELTIGMSLGTLDASCVELTINTSKDQASCTCGNKKTOELYYKLSDLGCFNFRIPPT 89
 QY 56 WSTHAYAFPPYKIMDRFFKRVDEVINGALKRGALVAIINHIEELMNDPEEHK-----DDEEHER 110
 DB 90 WSGHFGDAPDYKIDSVMMKRVHEVVDYALNTGCAIINH--ETWVAFQKNLESARKI 147
 QY 111 FLALMKQIADRYKDYPELFEELINEPHGNLTPEPK-----NELLEALKYIRS 159
 DB 148 LVALMKQIAEFQYDEHLIFEGNNEPKYVDPALWGTGQDEGNENYENMALFVKTIIRA 207
 QY 160 ID-----KKHTIIIGTAEMGISALEKLSVPKMEKNSIVTIIHYNPEFT--HOG-----AEW 210
 DB 208 TGGNANRHLMIPTAYASVNDGSIINFKYPRGDDKIVLSHSYSPYFALNNGCAISNF 267
 QY 211 VEGSE--KWLGRKWSDDQKHLIEEFNFIEEMSKKRRPIYIGEGAYRAADESRKWT 269
 DB 268 YDGNIEIDMV-----MNTINSSFTSKGIPVIGEGFVAMNRDNEDDRERWQ 311

Query Match 17.1% Score 296.5; DB 4; Length 471;
 Best Local Similarity 28.0%; Pred. No. 1e-20;
 Matches 98; Conservative 51; Mismatches 134; Indels 67; Gaps 14;

RESULT 15
 US-09-286-691-12
 Sequence 12, Application US/09286691
 Patent No. 6190189
 GENERAL INFORMATION:
 APPLICANT: Li, Xin-Liang
 APPLICANT: Ljungdahl, Lars G.
 APPLICANT: Chen, Huizhong
 TITLE OF INVENTION: Cellulases and Coding Sequences
 FILE REFERENCE: 42-96
 CURRENT APPLICATION NUMBER: US/09/286,691
 EARLIER FILING DATE: 1999-04-05
 EARLIER APPLICATION NUMBER: US 60/027,883
 EARLIER FILING DATE: 1996-10-04
 EARLIER APPLICATION NUMBER: PCT US97/18008
 EARLIER FILING DATE: 1997-10-03
 NUMBER OF SEQ ID NOS: 29
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 12
 LENGTH: 471
 TYPE: PRT
 ORGANISM: Orpinomyces sp. PC-2
 US-09-286-691-12

Query Match 17.1% Score 296.5; DB 4; Length 471;
 Best Local Similarity 28.0%; Pred. No. 1e-20;
 Matches 98; Conservative 51; Mismatches 134; Indels 67; Gaps 14;

QY 10 KILRGINIGNALAE-----PNECDWG--VVIDEFDIIEAGFSHVIRIPIR 55
 DB 30 KELTIGMSLGTLDASCVELTINTSKDQASCTCGNKKTOELYYKLSDLGCFNFRIPPT 89
 QY 56 WSTHAYAFPPYKIMDRFFKRVDEVINGALKRGALVAIINHIEELMNDPEEHK-----DDEEHER 110
 DB 90 WSGHFGDAPDYKIDSVMMKRVHEVVDYALNTGCAIINH--ETWVAFQKNLESARKI 147
 QY 111 FLALMKQIADRYKDYPELFEELINEPHGNLTPEPK-----NELLEALKYIRS 159
 DB 148 LVALMKQIAEFQYDEHLIFEGNNEPKYVDPALWGTGQDEGNENYENMALFVKTIIRA 207
 QY 160 ID-----KKHTIIIGTAEMGISALEKLSVPKMEKNSIVTIIHYNPEFT--HOG-----AEW 210
 DB 208 TGGNANRHLMIPTAYASVNDGSIINFKYPRGDDKIVLSHSYSPYFALNNGCAISNF 267
 QY 211 VEGSE--KWLGRKWSDDQKHLIEEFNFIEEMSKKRRPIYIGEGAYRAADESRKWT 269
 DB 268 YDGNIEIDMV-----MNTINSSFTSKGIPVIGEGFVAMNRDNEDDRERWQ 311

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2003, 13:59:59 ; Search time 20 Seconds
(without alignments)
1270.057 Million cell updates/sec

Title: US-09-880-729A-2

Perfect score: 1732

Sequence: 1 MGVDFPERKILGRGINICN.....LRKYNKLLLEALIGDSIE 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications-AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1732	100.0	317	9	US-09-880-729-2
2	835.5	48.2	360	9	US-10-121-032-27
3	306.5	17.7	574	10	US-09-863-547B-2
4	292	16.9	395	9	US-09-797-464A-2
5	285	16.5	400	9	US-09-797-464A-4
6	283.5	16.4	348	9	US-09-797-464A-7
7	279	16.1	363	9	US-09-797-464A-11
8	203	11.7	160	9	US-09-797-464A-8
9	203	11.7	167	9	US-09-797-464A-6
10	177	10.2	308	9	US-10-068-374-5
11	177	10.2	467	9	US-10-068-374-2
12	177	10.2	467	10	US-09-863-547B-1
13	168	9.7	400	9	US-09-903-185-2
14	168	9.7	462	9	US-09-903-185-13
15	159.5	9.2	335	9	US-09-927-827-52
16	147.5	8.5	1621	9	US-10-185-990-10
17	143.5	8.3	956	9	US-10-121-032-63
18	139.5	8.1	327	9	US-09-916-494A-16
19	138.5	8.0	333	9	US-09-927-827-51

20	136.5	7.9	535	9	US-09-927-827-50	Sequence 50, Appl
21	129	7.4	325	9	US-09-927-827-17	Sequence 17, Appl
22	128	7.4	427	9	US-09-850-982B-2	Sequence 2, Appl
23	123.5	7.1	553	10	US-09-888-224-2	Sequence 2, Appl
24	121.5	7.0	558	9	US-09-862-660-2	Sequence 2, Appl
25	121	7.0	663	9	US-10-121-032-61	Sequence 61, Appl
26	121	7.0	680	9	US-10-121-032-25	Sequence 25, Appl
27	116.5	6.7	454	9	US-10-121-032-20	Sequence 20, Appl
28	112.5	6.5	166	9	US-09-797-464A-5	Sequence 5, Appl
29	111.5	6.4	174	9	US-09-797-464A-9	Sequence 9, Appl
30	110.5	6.4	333	9	US-09-927-827-3	Sequence 3, Appl
31	108	6.2	562	10	US-09-981-900B-5	Sequence 5, Appl
32	105	6.1	464	10	US-09-815-242-10647	Sequence 10647, A
33	105	6.1	467	10	US-09-815-242-4997	Sequence 4997, Ap
34	96.5	5.6	423	10	US-09-369-735B-5	Sequence 5, Appl
35	96.5	5.6	423	10	US-09-369-735B-5	Sequence 5, Appl
36	95.5	5.5	329	9	US-09-927-827-16	Sequence 16, Appl
37	95	5.5	602	9	US-10-195-158-5	Sequence 5, Appl
38	93	5.4	421	9	US-10-121-032-17	Sequence 17, Appl
39	93	5.4	511	9	US-10-121-032-21	Sequence 21, Appl
40	92.5	5.3	515	10	US-09-765-272-22	Sequence 22, Appl
41	92.5	5.3	684	9	US-10-121-032-62	Sequence 62, Appl
42	91.5	5.3	459	10	US-09-815-242-13229	Sequence 13229, A
43	90.5	5.2	370	9	US-09-738-626-5196	Sequence 5196, Ap
44	90	5.2	489	10	US-09-369-735B-9	Sequence 9, Appl
45	87	5.0	438	9	US-10-121-032-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-880-729-2
Sequence 2, Application US/09880729
Publication No. US20030044956A1
GENERAL INFORMATION:
APPLICANT: Matchur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from
Thermotoga Maritima
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: CARELLA, BYRNE, BAIN, GILFILLAN,
CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/880,729
FILING DATE: 12-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/472,857
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/951,889
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 317 AMINO ACIDS
 TYPE: AMINO ACID
 STRANDEDNESS: <Unknown>
 TOPOLOGY: LINEAR
 MOLECULE TYPE: PROTEIN
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 us-09-880-729-2

Query Match 100.0%; Score 1732; DB 9; Length 317;
 Best Local Similarity 100.0%; Pred. No. 3,2e-146;
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGVDPPEKRIILGRGINIGNALAPNEGDMGVYIKDFEDIIKEAGFSVRIPIRSTHA 60
 DB 1 MGVDPPEKRIILGRGINIGNALAPNEGDMGVYIKDFEDIIKEAGFSVRIPIRSTHA 60
 QY 61 YAFPPYKIDREFKRYDEVINGALKRGLAVAINIHHEELMNDPEEKKERFLALMQIAD 120
 DB 61 YAFPPYKIDREFKRYDEVINGALKRGLAVAINIHHEELMNDPEEKKERFLALMQIAD 120
 QY 121 RYKDYPELTFEELNPHGNLTPPEKWNELLEALKVIRSIDKKHTIIIGTAEMGISALE 180
 DB 121 RYKDYPELTFEELNPHGNLTPPEKWNELLEALKVIRSIDKKHTIIIGTAEMGISALE 180
 QY 181 KISVPKWEKNSIVTIHYNPFEPFHOGAEWEGSEKMLGRKMGSPDDOKHLEEFNFIEE 240
 DB 181 KISVPKWEKNSIVTIHYNPFEPFHOGAEWEGSEKMLGRKMGSPDDOKHLEEFNFIEE 240
 QY 241 WSKNNRPYIYIGFGAYRRADLESRIKWTSPVVREREMKRRMSAWYEFCSGFGVDTLTK 300
 DB 241 WSKNNRPYIYIGFGAYRRADLESRIKWTSPVVREREMKRRMSAWYEFCSGFGVDTLTK 300
 QY 301 TWNKDLLEALLIGDSTE 317
 DB 301 TWNKDLLEALLIGDSTE 317

RESULT 2

US-10-121-032-27
 Sequence 27, Application US/10121032
 Patent No. US2002015550A1

GENERAL INFORMATION:
 APPLICANT: Bylina, Edward J.
 TITLE OF INVENTION: GLYCOSIDASE ENZYMES
 NUMBER OF SEQUENCES: 72
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Gray Cary Ware & Freidenrich LLP
 STREET: 4365 Executive Drive, Suite 1600
 CITY: San Diego
 STATE: CA
 COUNTRY: USA
 ZIP: 92121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows95
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/121.032
 FILING DATE: 09-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/134,078
 FILING DATE: 13-AUG-1998
 APPLICATION NUMBER: 08/949,026
 FILING DATE: 10-OCT-1997
 APPLICATION NUMBER: 60/056,916
 FILING DATE: 06-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Hallie, Lisa A.
 REGISTRATION NUMBER: 38,347
 REFERENCE/DOCKET NUMBER: 09010/024002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 858/677-1456
 TELEFAX: 858/677-1465
 INFORMATION FOR SEQ ID NO: 27:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 360 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: Internal
 SEQUENCE DESCRIPTION: SEQ ID NO: 27:
 us-10-121-032-27

Query Match 48.2%; Score 835.5; DB 9; Length 360;
 Best Local Similarity 49.5%; Pred. No. 2e-66;
 Matches 154; Conservative 53; Mismatches 95; Indels 9; Gaps 3;

QY 6 FERNKILGRGINIGNALAPNEGDMGVYIKDFEDIIKEAGFSVRIPIRSTHA 65
 DB 51 FEYKMWGKGVNIGNALAPNEGDMGVYIKDFEDIIKEAGFSVRIPIRSTHA 110
 QY 66 YKIMDFEKKRYDEVINGALKRGLAVAINIHHEELMNDPEEKKERFLALMQIADRYKDY 125
 DB 111 YDIDRNFLEKRVHVVDRALENNLIYINHHFEELXQEPDKGVYVLEIMRQIAKFEKDY 170
 QY 126 PETLFEELNPHGNLTPPEKWNELLEALKVIRSIDKKHTIIIGTAEMGISALEKISVP 185
 DB 171 PETLFEELNPHGNLTPPEKWNELLEALKVIRSIDKKHTIIIGTAEMGISALEKISVP 230
 QY 186 KMEKNSIVTIHYNPFEPFHOGAEWEGSEKMLGRKMGSPDDOKHLEEFNFIEE 241
 DB 231 N-DKRITVSFHYEPKRYHOGAPVNPVRYVKNGEWEIINDISH---FKYSDW 285
 QY 242 SKNNRPYIYIGFGAYRRADLESRIKWTSPVVREREMKRRMSAWYEFCSGFGVDTLTK 301
 DB 286 AKONNVPYIYIGFGAYRRADLESRIKWTSPVVREREMKRRMSAWYEFCSGFGVDTLTK 345
 QY 302 TWNKDLLEALLIG 312
 DB 346 WLEPLATAVVG 356

RESULT 3

US-09-863-547B-2
 Sequence 2, Application US/09863547B
 Patent No. US20020128166A1

GENERAL INFORMATION:
 APPLICANT: Henkel KGaA
 APPLICANT: Van Beekhoven, Lentus B. M.
 APPLICANT: Maurer, Karl-Heinz
 APPLICANT: Kottwitz, Beatrix

APPLICANT: Weis, Albrecht
 APPLICANT: Van Solingen, Pieter
 TITLE OF INVENTION: Detergents Comprising Cellulases
 FILE REFERENCE: H 1920 A

CURRENT APPLICATION NUMBER: US/09/863,547B
 CURRENT FILING DATE: 2002-02-13
 PRIOR APPLICATION NUMBER: US 08/945,574
 PRIOR FILING DATE: 1998-02-27

PRIOR APPLICATION NUMBER: PCT/EP96/01755
 PRIOR FILING DATE: 1996-04-26
 PRIOR APPLICATION NUMBER: US 08/614,115
 PRIOR FILING DATE: 1996-03-12
 PRIOR APPLICATION NUMBER: EP 95201115.3
 PRIOR FILING DATE: 1995-04-28
 NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn version 3.1
 SEQ ID NO 2
 LENGTH: 574
 TYPE: PRT

ORGANISM: Bacillus sp. 669.93
 US-09-863-547B-2


```
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 167
; TYPE: PRF
; ORGANISM: Paenibacillus pabuli
US-09-797-464A-6
```

```
Query Match
Best Local Similarity 11.7%; Score 203; DB 9; Length 167;
Matches 49; Conservative 25; Mismatches 50; Indels 16; Gaps 5;
```

```
QY 13 GRCIGNMALEA-----PNEGDMG-VVVKDEFDIIKEAGFSHVRIPIRMSHAYAPPPY 66
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 GAGNMGNGLEAAVNGPTNPTAMGNPTVPELIIKKVAKGKSLRIPVSLNIGSAPNY 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 67 KIDRFKRVDEVYNGKLKGLAVAINIH--HFE-----LMDPEEH--KERFLAKM 116
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 TINAAWLRIOQVVDYVYNGELVYINIHGDYNSVGGMLLVNGNQTAKIKKKYKQW 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 117 QIADRYKDYPETLEFFELINE 136
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 QIATKFSNYMDRLIFESMNE 140
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 10
US-10-068-374-5
; Sequence 5, Application US/10068374
; Publication No. US20030044903A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Shaw, Andrew
; TITLE OF INVENTION: Modified Target Enzymes Having Catalytic
; TITLE OF INVENTION: Triads
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: GC652
; CURRENT APPLICATION NUMBER: US/10/068,374
; CURRENT FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: US 60/268,347
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 308
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: protein encoded by plasmid pCORE3
US-10-068-374-5
```

```
Query Match
Best Local Similarity 10.2%; Score 177; DB 9; Length 308;
Matches 67; Conservative 40; Mismatches 94; Indels 64; Gaps 15;
```

```
QY 29 DWGVVVKDEFDDIIKEAGFSHVRIPIRMSHAYAPPPYKIMDRFKRVDEVYNGALKRGL 88
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 55 DWGITV-----FRAMYTSSGGYIDP-----SVKEKVKETVEAIDLGI 94
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 89 AVAINIHHEELMNDPEEHKERFLAKMQLADRYKDYPETLEFFELINEPHGNLTPEKWE 148
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 95 YVILDMHILSD--NDPNVYKEAKDFDEMSGLXGDPNNV-YEIAEPNG--SDVTWDN 149
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 149 LL-----EALAKVIRSIDKKHTIIGTAEWG---GISALEKLSVPKWKNSIVTTHYNNP 201
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 150 QIKRYAEVIVPIRDNDPNVIVIGTWSODVHHAADNOLADP-----NVMYAFHFY--- 202
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 202 EFTHOGAEWEGSEKWKGRKWSGPDOKHLIEENFLEWSKKRRPIYIGEGGAYRKAD 261
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 203 AGTH-GONLRDQVYAL-----DOGAAL-----FVSEWGTG---AATGDDGVF----- 241
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 262 LESRIKWTSPVVREREMKRRMSWAYW 286
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 242 LDEAQVWIDF-----MDERNLSMANW 262
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 11
US-10-068-374-2
; Sequence 2, Application US/10068374
; Publication No. US20030044903A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Shaw, Andrew
; TITLE OF INVENTION: Modified Target Enzymes Having Catalytic
; TITLE OF INVENTION: Triads
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: GC652
; CURRENT APPLICATION NUMBER: US/10/068,374
; CURRENT FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: US 60/268,347
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 467
; TYPE: PRF
; ORGANISM: Bacillus sp.
US-10-068-374-2
```

```
Query Match
Best Local Similarity 10.2%; Score 177; DB 9; Length 467;
Matches 67; Conservative 40; Mismatches 94; Indels 64; Gaps 15;
```

```
QY 29 DWGVVVKDEFDDIIKEAGFSHVRIPIRMSHAYAPPPYKIMDRFKRVDEVYNGALKRGL 88
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 81 DWGITV-----FRAMYTSSGGYIDP-----SVKEKVKETVEAIDLGI 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 89 AVAINIHHEELMNDPEEHKERFLAKMQLADRYKDYPETLEFFELINEPHGNLTPEKWE 148
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 YVILDMHILSD--NDPNVYKEAKDFDEMSGLXGDPNNV-YEIAEPNG--SDVTWDN 175
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 149 LL-----EALAKVIRSIDKKHTIIGTAEWG---GISALEKLSVPKWKNSIVTTHYNNP 201
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 176 QIKRYAEVIVPIRDNDPNVIVIGTWSODVHHAADNOLADP-----NVMYAFHFY--- 228
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 202 EFTHOGAEWEGSEKWKGRKWSGPDOKHLIEENFLEWSKKRRPIYIGEGGAYRKAD 261
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 229 AGTH-GONLRDQVYAL-----DOGAAL-----FVSEWGTG---AATGDDGVF----- 267
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 262 LESRIKWTSPVVREREMKRRMSWAYW 286
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 268 LDEAQVWIDF-----MDERNLSMANW 288
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 12
```

```
US-09-863-547B-1
; Sequence 1, Application US/09863547B
; Patent No. US20020128166A1
; GENERAL INFORMATION:
```

SOFTWARE: Patentln version 3.1
SEQ ID NO 1
LENGTH: 467
TYPE: PRT
ORGANISM: Bacillus sp. CBS 670.93
US-09-863-547B-1

Query Match 10.2%; Score 177; DB 10; Length 467;
Best Local Similarity 25.3%; Pred. No. 1e-07;
Matches 67; Conservative 40; Mismatches 94; Indels 64; Gaps 15;

QY 29 DWGVYIKDFEFDIIEAGFSHVRIPIRMTSHAYAFPPYKIMDRFFKRVDEVINGALRGRL 88
DB 81 DWGITV-----FRAAMVYSSGGYIDDP-----SVKEKVEYEAALDLDI 120
QY 89 AVAINIHHEELNDPEEKERFLALMKQIADRYKDPETLEFEILNEPFGNLTPKRWNE 148
DB 121 VYIIDMHILSD--NDPNITKEEAKDFDEMSELYGDPNVI-YEIANEPNG--SDVTWGN 175
QY 149 LL-----EALKYIRSIDKHTIIGTAENG--GISALEKLSVPKWEKNSIVTIHYNPF 201
DB 176 QIKPYAEVYIPVIRNDPNNTIIVGTGTSQDVHHAADNQLADP---NVMYAFHEY--- 228
QY 202 EFTHOGAEWESESEKVLGRKWSRDDQKHLIEEFNFIEEMSKKRRPIYIGEGAYRKAD 261
DB 229 AGTH-GQNLRDQVDYAL-----DQGAAT---FVSEWGTG---AATGDGCVF----- 267
QY 262 LESRIKTSFVYVREMEKRRMSWAYW 286
DB 268 LDEAQYWIDF---MDERNLSWANW 288

RESULT 13
US-09-903-185-2
Sequence 2, Application US/09903185
Publication No. US20030054534A1
GENERAL INFORMATION:

APPLICANT: Outtrup, Helle
Dandmann, Claus
Olsen, Arne
Bisq Id-Frantzen, Henrik
Sch Jeln, Martin
Jrgensen, Per
Bjoernvad, Mads
TITLE OF INVENTION: DNA Constructs and Methods of Producing
Cellulytic Enzymes
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: NO. US20030054534A10 NO. US20030054534A1disk of No. US20030054534A1
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/903,185
FILING DATE: 11-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/870,180
FILING DATE: 6-June-1997
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 3794.434-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-903-185-2

Query Match 9.7%; Score 168; DB 9; Length 400;
Best Local Similarity 24.9%; Pred. No. 5.1e-07;
Matches 66; Conservative 43; Mismatches 92; Indels 64; Gaps 16;

QY 29 DWGVYIKDFEFDIIEAGFSHVRIPIRMTSHAYAFPPYKIMDRFFKRVDEVINGALRGRL 88
DB 81 DWGITV-----NVEFAAMT-----SSGGYIDDP-----SVKEKVEYEAALDLDI 120
QY 89 AVAINIHHEELNDPEEKERFLALMKQIADRYKDPETLEFEILNEPFGNLTPKRWNE 148
DB 121 VYIIDMHILSD--NDPNITKEEAKDFDEMSELYGDPNVI-YEIANEPNG--SDVTWGN 175
QY 149 LL-----EALKYIRSIDKHTIIGTAENG--GISALEKLSVPKWEKNSIVTIHYNPF 201
DB 176 QIKPYAEVYIPVIRNDPNNTIIVGTGTSQDVHHAADNQLADP---NVMYAFHEY--- 228
QY 202 EFTHOGAEWESESEKVLGRKWSRDDQKHLIEEFNFIEEMSKKRRPIYIGEGAYRKAD 261
DB 229 AGTH-GQNLRDQVDYAL-----DQGAAT---FVSEWGTG---AATGDGCVF----- 267
QY 262 LESRIKTSFVYVREMEKRRMSWAYW 286
DB 268 LDEAQYWIDF---MDERNLSWANW 288

RESULT 14
US-09-903-185-13
Sequence 13, Application US/09903185
Publication No. US20030054534A1
GENERAL INFORMATION:

APPLICANT: Outtrup, Helle
Dandmann, Claus
Olsen, Arne
Bisq Id-Frantzen, Henrik
Sch Jeln, Martin
Jrgensen, Per
Bjoernvad, Mads
TITLE OF INVENTION: DNA Constructs and Methods of Producing
Cellulytic Enzymes
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: NO. US20030054534A10 NO. US20030054534A1disk of No. US20030054534A1
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/903,185
FILING DATE: 11-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/870,180
FILING DATE: 6-June-1997
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 3794.434-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298

TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-903-185-13

Query Match 9.7%; Score 168; DB 9; Length 462;
Best Local Similarity 24.9%; Pred. No. 6.2e-07;
Matches 66; Conservative 43; Mismatches 92; Indels 64; Gaps 16;

QY 29 DMGVVIRKDEFFDIKEAGFSHVIRPIRMSTHAYAFPPKIMDRFKEFVDEVLNGALRGRL 88
DB 81 DWGI-----NVFRAMYT-----SSGCIYDDP-----SVKEKVEAVERAIDLDT 120
QY 89 AVAINIHVEELNNDPEHKEERFLALMKQIADRYKDYPTLFEELNPEHNLTPERKNE 148
DB 121 YVLIIDWHILSD--NDPRITYKEFAKDFEDMSELYGOVPIV-YEIANEPNG--SDVTWGN 175
QY 149 LL-----EALKVIRSIDKKHTIIIGTAEWG--GISALEKLSVPKWEKNSIVTIHYNPF 201
DB 176 QIRPYAEVPIIRNDPNIIIVGTGTSQDVHHAADNOLADP---NMYAFHFY--- 228
QY 202 EFTHOGAEWEGSEKMGGRKWSPPDOCKHLEEFNFIEEWSKKKRPRIYGEFGAYRKAD 261
DB 229 AGTH-GONLRDOVDYAL-----DGAALF---FVSEWGTG---AATGDGVP--- 267
QY 262 LESRIKWTSEVREMERKRWMAWY 286
DB 268 LDEAQWIDF---MDERNLSWANN 288

RESULT 15
US-09-927-827-52
Sequence 52, Application US/09927827
Publication No. US20030036176A1
GENERAL INFORMATION:
APPLICANT: Bower, Stanley G.
APPLICANT: Ramsdell, Thomas M.
TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
FILE REFERENCE: 38-10(15824)B
CURRENT APPLICATION NUMBER: US/09/927,827
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: US 60/279,493
PRIOR FILING DATE: 2001-03-28
NUMBER OF SEQ ID NOS: 69
SEQ ID NO 52
LENGTH: 335
TYPE: PRT
ORGANISM: Xanthomonas campestris
US-09-927-827-52

Query Match 9.2%; Score 159.5; DB 9; Length 335;
Best Local Similarity 21.0%; Pred. No. 2.3e-06;
Matches 69; Conservative 55; Mismatches 133; Indels 71; Gaps 16;

QY 15 GINIGNALEAPNDEMGVIRKDEFFDIKE-----AGFSHVIRPIRMSTHAYAF---FP 64
DB 23 GVNLAGA-EFASSKPGVLNKKDYMPASTDYSEFAGTGMVIRLPILMERLOPAANGELD 81
QY 65 PYKIMDRFPRVDEVINGALRGGLAVAINIHVEEL---MNDPEHKEERFLALMKQIAD 120
DB 82 PAOL-----ALVOQAAVARAKASGMVYLVLDIHNSKYYGYKMGPEVPLATFADLMRLAV 136
QY 121 RYKDYPTLFEELNPEHNLTPERKNELEEFALKVIRSIDKKHTIIIGTAEWG----- 175
DB 137 IFNS-DNAVITFGLMERN-NISASEWAGAQAQIDAIRATGANNLLIVPGALMTGAHSMH 194
QY 176 -----ISALEKLSVPKWEKNSIVTIHYNPFETHOGAEWEGSEKMGGRKWSPPD 226

DB 195 SLTSDGYSNATALASISDPL--NRFAFEVHQY--LDADSSGTSVCVSET-IG----- 242
QY 227 DQKHLEEFNFIEEWSKKKRPRIYGEFGAYRKADLESRIKWTSEVREMERKRR----- 281
DB 243 -----ADRLRAFTEWLTNNKRGFLGEFGTANNVCFNALOG---MLAYMENYADVWLGW 294
QY 282 SW---AYWEEFCGFGVYDTLRKTWNKD 305
DB 295 TWMAGAMWNTSYAYNVHP-----NKD 316

Search completed: April 22, 2003, 14:05:39
Job time : 22 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 10:49:07 ; Search time 191 Seconds
(without alignments)
11248.201 Million cell updates/sec

Title: US-09-880-729A-1

Perfect score: 954
Sequence: 1 atgggtgtgtagctcttga.....gagagatagcatgataa 954

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues.

Word size : 20

Total number of hits satisfying chosen parameters: 23

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N.Geneseq_101002.*

```
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	954	100.0	954	AAZ21563	Carboxymethyl cell
2	954	100.0	954	AAZ21563	cDNA encoding a ca
3	954	100.0	954	AAZ46759	T. maritima CMCase
4	954	100.0	954	AAZ15243	Thermotoga maritim
5	21	2.2	51	AAZ21564	PCR primer for the
6	21	2.2	51	AAZ85860	PCR primer used to
7	21	2.2	51	AAZ46760	T. maritima CMCase
8	21	2.2	51	AAZ15244	Thermotoga maritim
9	21	2.2	530	AAZ35482	Human musculoskele

C 10	20	2.1	849	13	AAO32472	HCV NS2-NS4 clone
C 11	20	2.1	849	13	AAO32483	HCV NS2-NS4 clone
C 12	20	2.1	849	13	AAO32482	HCV NS2-NS4 clone
C 13	20	2.1	849	13	AAO32484	HCV NS2-NS4 clone
C 14	20	2.1	1280	13	AAO32479	HCV NS2-NS4 clone
C 15	20	2.1	1280	13	AAO32494	HCV NS2-NS4 clone
C 16	20	2.1	1280	13	AAO32495	HCV NS2-NS4 clone
C 17	20	2.1	1350	22	AAZ37283	Human musculoskele
C 18	20	2.1	1350	22	AAZ37284	Human musculoskele
C 19	20	2.1	1350	22	AAZ37285	Human musculoskele
C 20	20	2.1	3564	13	AAO32442	HCV NS2-NS4 clone
C 21	20	2.1	3564	13	AAO32501	HCV NS2-NS4 clone
C 22	20	2.1	4235	23	ABL28818	Drosophila melanog
C 23	20	2.1	7911	13	AAO32436	HCV antigen clone

ALIGNMENTS

```
RESULT 1
ID AAZ21563 standard: cDNA: 954 BP.
AC AAZ21563;
XX 01-DEC-1999 (first entry)
XX Carboxymethyl cellulase (CMCase) gene.
DE
XX Carboxymethyl cellulase; CMCase; thermostable enzyme; hydrolysis;
XX beta 1,4 glycosidic bond; cellulose; plant biomass conversion; fuel;
XX chemical; detergent; textile industry; animal feed; waste treatment;
XX fruit juice extraction; clarification; brewing industry; ss.
OS Thermotoga maritima.
FH
FT 1.954 Location/Qualifiers
FT CDS /tag= a
FT /product= CMCase
FT /note= "Carboxymethyl cellulase"
XX
XX US962258-A.
XX 05-OCT-1999.
XX
XX 23-AUG-1995; 950S-0518615.
XX
XX 23-AUG-1995; 950S-0518615.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Mathur EJ, Lam DE;
XX WPI: 1999-571269/48.
XX P-PSDB: AAZ39338.
XX
XX Recombinant polynucleotide sequence encoding a carboxymethyl cellulase
XX useful for the conversion of plant biomass -
XX
XX Claim 11: Fig 1; 14pp: English.
XX
XX This sequence is a carboxymethyl cellulase gene (CMCase). CMCase is a
XX thermostable enzyme (optimum temperature of 60 + degrees C), that
XX catalyzes the hydrolysis of the beta 1,4 glycosidic bonds in cellulose.
XX The enzyme degrades cellulose by the hydrolysis of this bond. The CMCase
XX enzyme has a molecular weight of approximately 35kd. CMCase is used to
XX catalyze enzymatic degradation of cellulose and can be used for the
XX conversion of plant biomass into fuels and chemicals. CMCase can also be
XX used in detergents, the textile industry, animal feed, waste treatment
XX and for the clarification and extraction of juices in the fruit juice and
XX brewing industries. The CMCase gene and 15 bp fragments of it can be used
XX as probes to isolate nucleic acid molecules encoding CMCase enzymes from
```

CC genomic libraries or to screen for related sequences. Cells expressing
 CC the CMCase polynucleotide can be used as immunogens to produce antibodies
 CC which can then be used to isolate the enzyme from cells which express
 CC CMCase. The antibodies can also be used to screen for similar enzymes
 CC from other organisms and samples.

XX Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other:

Query Match 100.0%; Score 954; DB 20; Length 954;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGGGTGTTGATCCTTTTGAAGAACAAATATTGGCAGAGCGATTATATAGGAAT 60
Db 1 ATGGGTGTTGATCCTTTTGAAGAACAAATATTGGCAGAGCGATTATATAGGAAT 60
QY 61 GCGCTTGAAGACCAATAGAGGAGAGCTGGGAGTGTGATTAAGATGAGTCTTCGAC 120
Db 61 GCGCTTGAAGACCAATAGAGGAGAGCTGGGAGTGTGATTAAGATGAGTCTTCGAC 120
QY 121 ATTATTAAGAGAGCGGTTCTCTCATGTTGCAATTCATTAAGATGAGTACGACGCT 180
Db 121 ATTATTAAGAGAGCGGTTCTCTCATGTTGCAATTCATTAAGATGAGTACGACGCT 180
QY 181 TACGGGTTCCCTTATTAATCATGATCGCTCTCTCAAAAGAGTGGATGAAGTATA 240
Db 181 TACGGGTTCCCTTATTAATCATGATCGCTCTCTCAAAAGAGTGGATGAAGTATA 240
QY 181 TACGGGTTCCCTTATTAATCATGATCGCTCTCTCAAAAGAGTGGATGAAGTATA 240
Db 181 TACGGGTTCCCTTATTAATCATGATCGCTCTCTCAAAAGAGTGGATGAAGTATA 240
QY 241 AACGAGCCCTGAAAGAGAGAGTGGCTGTGCTATTAATATTCATCATGAGAGATTA 300
Db 241 AACGAGCCCTGAAAGAGAGAGTGGCTGTGCTATTAATATTCATCATGAGAGATTA 300
QY 301 ATGAATGATCCAGAAAGACACAAAGAAATTTCTGCTCTTGGAAACAAATTTGCTGAT 360
Db 301 ATGAATGATCCAGAAAGACACAAAGAAATTTCTGCTCTTGGAAACAAATTTGCTGAT 360
QY 361 CCTTATTAAGATCCCAAACTCTATTTTGAATTCGATACATCACTCAGCAAAAT 420
Db 361 CCTTATTAAGATCCCAAACTCTATTTTGAATTCGATACATCACTCAGCAAAAT 420
QY 421 CTCTACTCCGGAATAATGATGACTGCTTGGAGAACTCTAAATTTATAGATCAAT 480
Db 421 CTCTACTCCGGAATAATGATGACTGCTTGGAGAACTCTAAATTTATAGATCAAT 480
QY 481 GACAAAGAACACACTATTAATTAATAGCAGACGATGAGTGGGGGTATCTGCTTGA 540
Db 481 GACAAAGAACACACTATTAATTAATAGCAGACGATGAGTGGGGGTATCTGCTTGA 540
QY 541 AAACGTCTGTCCCAAAATGGGAAAAAATCTATAGTTACAATTCACACTACATCTCT 600
Db 541 AAACGTCTGTCCCAAAATGGGAAAAAATCTATAGTTACAATTCACACTACATCTCT 600
QY 601 TTTCGAATTTACCATCAAGAGAGTGTGGTGGAGAGATCTGAGAAATGGTGGGAGA 660
Db 601 TTTCGAATTTACCATCAAGAGAGTGTGGTGGAGAGATCTGAGAAATGGTGGGAGA 660
QY 661 AAGTGGGAGTCTCAGATGATCAGAAACATTTAGTAGAAGAAATTCATTTATAGAGAA 720
Db 661 AAGTGGGAGTCTCAGATGATCAGAAACATTTAGTAGAAGAAATTCATTTATAGAGAA 720
QY 721 TGGTCAAAAAAGAACAAAGACCAATTTACATAGTGTGAGTGGCTTCAGAGAAAGCT 780
Db 721 TGGTCAAAAAAGAACAAAGACCAATTTACATAGTGTGAGTGGCTTCAGAGAAAGCT 780
QY 781 GACCTTGAATCAAGAAATTAATGAGACTCTTGTGCTGTGGCAATGGAGAAAGAGAGA 840
Db 781 GACCTTGAATCAAGAAATTAATGAGACTCTTGTGCTGTGGCAATGGAGAAAGAGAGA 840
QY 841 TGGAGCTGGGCATACCTGGAATTTTGTTCGAGTTTGGTGTATGATCTCTGAGAAAA 900
Db 841 TGGAGCTGGGCATACCTGGAATTTTGTTCGAGTTTGGTGTATGATCTCTGAGAAAA 900
QY 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAATAGAGAGATGACATTGAATA 954
  
```

Db 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAATAGAGAGATGACATTGAATA 954

RESULT 2

ID: AAX85879 standard; cDNA; 954 BP.

AXX85879;

10-SEP-1999 (first entry)

cDNA encoding a carboxymethyl cellulase enzyme.

Carboxymethyl cellulase enzyme; hydrolysis; cellulose; plant biomass;

Fuel; Chemical; detergent; textile industry; animal feed;

waste treatment; fruit juice; brewing industry; clarification;

juice extraction; ss.

Thermotoga maritima.

US925749-A.

20-JUL-1999.

23-AUG-1995; 95US-0518615.

23-AUG-1995; 95US-0518615.

24-APR-1998; 98US-0066075.

(DIVE-) DIVERSA CORP.

Lam DE, Mathur EJ;

MP1: 1999-429058/36.

P-PSDB; AAY23764.

New carboxymethyl cellulase enzyme is useful in cellulose digestion

Claim 1; Fig 1A-C; 13pp; English.

The present sequence encodes a carboxymethyl cellulase enzyme from

Thermotoga maritima. Oligonucleotide probes derived from the

polynucleotide sequence can be used to screen a library of genomic

DNA for recovery of the polynucleotide, as PCR primers to synthesize

full length polynucleotides or as diagnostic probes useful for

screening for similar enzymes from other organisms and samples.

The carboxymethyl cellulase enzymes can be used to catalyze hydrolysis

of cellulose for the conversion of plant biomass into fuels and

chemicals, for use in detergents, the textile industry, in animal feed,

waste treatment and in the fruit juice/brewing industry for the

clarification and extraction of juices.

Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other:

Query Match 100.0%; Score 954; DB 20; Length 954;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
Db 181 TACGCGTTTCCTTATATAAATCATGATCGCTTCTTCAAAAAGTGTAGTATGATA 240
QY 241 AACGAGCCCTGAAAAAGAGACTGGCTGTGATATAATATTCATCATGAGAGATTGA 300
Db 241 AACGAGCCCTGAAAAAGAGACTGGCTGTGATATAATATTCATCATGAGAGATTGA 300
QY 301 ATGAATGATCCAGAGACACAAAGAAAGATTCTTGTCTTTGGAAACAAATTTGCTGAT 360
Db 301 ATGAATGATCCAGAGACACAAAGAAAGATTCTTGTCTTTGGAAACAAATTTGCTGAT 360
QY 361 CGTTTAAAGACTATCCCGAAACTCTATTTTGTGAATTTGTGATGACCTCACGGAAT 420
Db 361 CGTTTAAAGACTATCCCGAAACTCTATTTTGTGAATTTGTGATGACCTCACGGAAT 420
QY 421 CTGACCTCCGAAAAATGATGATGACCTGTGAGAAAGCTTAAAGTTATAGTCAATT 480
Db 421 CTGACCTCCGAAAAATGATGATGACCTGTGAGAAAGCTTAAAGTTATAGTCAATT 480
QY 481 GACAAAAAGCACACTATATATATAGGCACAGCTGAATGGGGGGTATATCTGCCCTTGA 540
Db 481 GACAAAAAGCACACTATATATATAGGCACAGCTGAATGGGGGGTATATCTGCCCTTGA 540
QY 541 AAATGCTGTCTCCCAAAATGAGAAAAAATTTCTATAGTTACATTCACCTACTCAATCCT 600
Db 541 AAATGCTGTCTCCCAAAATGAGAAAAAATTTCTATAGTTACATTCACCTACTCAATCCT 600
QY 601 TTGGAATTTACCATCAGAGAGCTGATGGGTGGAAGAGATCTAGAAATGTTGGGAGA 660
Db 601 TTGGAATTTACCATCAGAGAGCTGATGGGTGGAAGAGATCTAGAAATGTTGGGAGA 660
QY 661 AAGTGGGGATCTCCAGATGATCAGAAAACATTTGATAGAAAGATTTTATAGAAAG 720
Db 661 AAGTGGGGATCTCCAGATGATCAGAAAACATTTGATAGAAAGATTTTATAGAAAG 720
QY 721 TGGTCAAAAAAGAACAAAGCAATTTACATAGGTGATGTTGCTGCTTACACAAAAGCT 780
Db 721 TGGTCAAAAAAGAACAAAGCAATTTACATAGGTGATGTTGCTGCTTACACAAAAGCT 780
QY 781 GACCTGGAATCAAGATTAATAATGACCTCTTGTGTGCGGAAATGAGAAAGAGAGA 840
Db 781 GACCTGGAATCAAGATTAATAATGACCTCTTGTGTGCGGAAATGAGAAAGAGAGA 840
QY 841 TGGAGCTGGGCATAGTGGGAATTTTGTCCGTTTGGTGTATATGATCTGAGAAAA 900
Db 841 TGGAGCTGGGCATAGTGGGAATTTTGTCCGTTTGGTGTATGATCTGAGAAAA 900
QY 901 AACTGGAATTAAGATCTTTAGAAAGCTTTATAGAGAGAGATAGCATGTGATAA 954
Db 901 AACTGGAATTAAGATCTTTAGAAAGCTTTATAGAGAGAGATAGCATGTGATAA 954

RESULT 3
AAZ46759
ID AAZ46759 standard; cDNA: 954 BP.
XX
AC AAZ46759;
XX
DT 31-MAR-2000 (first entry)
XX
DE T. maritima CMCase enzyme encoding cDNA.
XX
KW Endoglucanase; carboxymethyl cellulase; cellulose; CMCase; plant biomass;
KW beta-1,4-glycosidic bond; detergent; textile; guar gum; animal feed;
KW waste product; fuel; brewing; thermostable; ss.
XX
OS Thermotoga maritima.
XX
FH Key Location/Qualifiers
FT CDS 1..954
FT /tag= a
FT /product= "CMCase"
XX
PN US6008032-A.
```

```
XX
PD 28-DEC-1999.
XX
PF 16-OCT-1997; 97US-0951889.
XX
PR 23-AUG-1995; 95US-0518615.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Mathur EJ, Lam DE;
DR WPI: 2000-105335/09.
DR P-PSDB: AAY56814.
XX
PT Endoglucanase with carboxymethyl cellulase activity from Thermotoga
PT maritima, useful for degrading biomass to fuels and chemicals
XX
PS Disclosure: Fig 1A-C; 14pp; English.
XX
CC This cDNA encodes a T. maritima endoglucanase with carboxymethyl
CC cellulase activity. The enzyme CMCase is thermostable and used to degrade
CC carboxymethyl cellulose (CMC) or more generally any oligosaccharide that
CC contains beta-1,4-glycosidic bonds, particularly in detergents, textiles,
CC guar gum, animal feed, plant biomass or waste products. The CMCase enzyme
CC is especially used to convert plant biomass to chemicals and fuels, and
CC also for clarification and extraction of juice (e.g. in the brewing
CC industry). CMCase can also be used to raise specific antibodies.
XX
SQ Sequence 954 BP; 335 A; 145 C; 221 G; 253 T; 0 other;

Query Match 100.0%; Score 954; DB 21; Length 954;
Best local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCTTGATCCTTTTGAAGAGACAAATATTTGGGAAGAGGCAATTAATATAGCAAT 60
Db 1 ATGGGCTTGATCCTTTTGAAGAGACAAATATTTGGGAAGAGGCAATTAATATAGCAAT 60
QY 61 GCCCTGAACACCAATGAGGGAGCTGGGAGTGGTATTAAGATGATCTTGTGAC 120
Db 61 GCCCTGAACACCAATGAGGGAGCTGGGAGTGGTATTAAGATGATCTTGTGAC 120
QY 121 ATTATAAAGAGCCGGTTTCTCATGTTGCAATTCGAATTAAGATGATGAGTACGACGCT 180
Db 121 ATTATAAAGAGCCGGTTTCTCATGTTGCAATTCGAATTAAGATGATGAGTACGACGCT 180
QY 181 TACGCTTTCTCTTATTAATATCATGATCGCTTCTTCAAAAGAGTGTGATGATGATA 240
Db 181 TACGCTTTCTCTTATTAATATCATGATCGCTTCTTCAAAAGAGTGTGATGATGATA 240
QY 241 AACGAGCCCTGAAAAAGAGAGTGGCTGTGATATAATATTCATCATGAGAGATTGA 300
Db 241 AACGAGCCCTGAAAAAGAGAGTGGCTGTGATATAATATTCATCATGAGAGATTGA 300
QY 301 ATGAATGATCCAGAGACACAAAGAAAGATTCTTGTCTTTGGAAACAAATTTGCTGAT 360
Db 301 ATGAATGATCCAGAGACACAAAGAAAGATTCTTGTCTTTGGAAACAAATTTGCTGAT 360
QY 361 CGTTTAAAGACTATCCCGAAACTCTATTTTGTGAATTTGTGATGACCTCACGGAAT 420
Db 361 CGTTTAAAGACTATCCCGAAACTCTATTTTGTGAATTTGTGATGACCTCACGGAAT 420
QY 421 CTGACCTCCGAAAAATGATGATGACCTGTGAGAAAGCTTAAAGTTATAGTCAATT 480
Db 421 CTGACCTCCGAAAAATGATGATGACCTGTGAGAAAGCTTAAAGTTATAGTCAATT 480
QY 481 GACAAAAAGCACACTATATATATAGGCACAGCTGAATGGGGGGTATATCTGCCCTTGA 540
Db 481 GACAAAAAGCACACTATATATATAGGCACAGCTGAATGGGGGGTATATCTGCCCTTGA 540
QY 541 AAATGCTGTCTCCCAAAATGAGAAAAAATTTCTATAGTTACATTCACCTACTCAATCCT 600
Db 541 AAATGCTGTCTCCCAAAATGAGAAAAAATTTCTATAGTTACATTCACCTACTCAATCCT 600
```

QY	601	TTCCAAATTTTACCACCATCAAGGAGCTGAGTGGGTGCAGAGATCTCGAANAATGCTTGGGAAGA	660
Db	601	TTTCGAATTTTACCACCATCAAGGAGCTGAGTGGGTGCAGAGATCTCGAANAATGCTTGGGAAGA	660
QY	661	AAGTGGGGATCTCCAGATGATCAGAACATTGATAGAACAAATTTCAATTTTATAGAGAA	720
Db	661	AAGTGGGGATCTCCAGATGATCAGAACAAATTTGATAGAACAAATTTCAATTTTATAGAGAA	720
QY	721	TGTCACAAAAAGACAACAAAGACACATTTTACATAGTAGTGTGGTGCTCACAGAAAAGCT	780
Db	721	TGTCACAAAAAGACAACAAAGACACATTTTACATAGTAGTGTGGTGCTCACAGAAAAGCT	780
QY	781	GACCTGGAATCAAAATAATTAATGAGCTCTCTTTGCTGTCGCCAAATGAGAAAAAGGAGA	840
Db	781	GACCTGGAATCAAAATAATTAATGAGCTCTCTTTGCTGTCGCCAAATGAGAAAAAGGAGA	840
QY	841	TGGACCTGGGCATACTGGGAATTTTGTTCGCGTTTTGGTGTGTTGATPACTCTGAGAAA	900
Db	841	TGGACCTGGGCATACTGGGAATTTTGTTCGCGTTTTGGTGTGTTGATPACTCTGAGAAA	900
QY	901	ACCTGGAATTAAGATCTTTTGAAGCTTTTAATAGAGAGATAGCATTTGAATAA	954
Db	901	ACCTGGAATTAAGATCTTTTGAAGCTTTTAATAGAGAGATAGCATTTGAATAA	954
 RESULT 4 AAD15243 ID AAD15243 standard; DNA; 954 BP.			
XX	AC	AAD15243;	
XX	DT	15-NOV-2001 (first entry)	
XX	DE	Thermotoga maritima CMCcase enzyme encoding DNA.	
KW	KM	Carboxymethyl cellulose; CMCcase; plant biomass; animal feed; detergent; textile industry; waste treatment; brewing industry; immunogen;	
KW	KM	fruit juice industry; ds.	
OS	XX	Thermotoga maritima.	
XX	FH	Key Location/Qualifiers	
FT	FT	CDS 1..954	
FT	FT	/tag= "a	
FT	FT	/product= "T. maritima CMCcase enzyme "	
PN	XX	US6245547-BI.	
PD	XX	12-JUN-2001.	
PF	XX	27-DEC-1999; 99US-0472857.	
PR	XX	23-AUG-1995; 95US-0518615.	
PR	XX	16-OCT-1997; 97US-0951889.	
PA	XX	(DIVE-) DIVERSA CORP.	
PI	XX	Mathur ET, Lam DE;	
DR	XX	WPI; 2001-528380/58.	
DR	XX	P-PDB; AA08543.	
PT	XX	Polynucleotide isolated from thermophile eubacteria useful for	
PT	XX	degradation of cellulose encodes an enzyme having a fully defined amino	
PS	XX	acid sequence -	
XX	XX	Claim 3; Fig 1; 14pp; English.	
CC	CC	The present sequence is a DNA coding for a thermostable enzyme, an	
CC	CC	endo glucanase having carboxymethyl cellulose activity referred to as	
CC	CC	(CMCase). The CMCase enzyme is useful for the degradation of cellulose,	
CC	CC	for conversion of plant biomass into fuels and chemicals, in the	

Query Match	Best Local Similarity	Score 954:	DB 22:	Length 954:
Matches 954:	Conservative	0:	Mismatches	0:
			Indels	Gaps
Qy 1	ATGGGTGTTGATCCTTTTGAAGGAACAATAATTTGGAGAGAGCATTATATAGAAAT	60		
Db 1	ATGGGTGTTGATCCTTTTGAAGGAACAATAATTTGGAGAGCATTATATAGAAAT	60		
Qy 61	GGCGTTGAAGCACAATAATGAGGAGACTGGGGAGTGGCTATTAAGAATGATTTCTTGAC	120		
Db 61	GGCGTTGAAGCACAATAATGAGGAGACTGGGGAGTGGCTATTAAGAATGATTTCTTGAC	120		
Qy 121	ATTATATAAAGAAAGCGGTTTCTCATGTTCCGATTCACATTAAGATGAGTTCGCACGCT	180		
Db 121	ATTATATAAAGAAAGCGGTTTCTCATGTTCCGATTCACATTAAGATGAGTTCGCACGCT	180		
Qy 181	TACGGCTTTCCTCCTTATTAATATGATGATCGCTTCTTAAAGAGCTGATGAGATGATA	240		
Db 181	TACGGCTTTCCTCCTTATTAATATGATGATCGCTTCTTAAAGAGCTGATGAGATGATA	240		
Qy 241	AACGGAGCCCTTAANAAGAGAGCTGGCTGGTGCATTAATAATTATCATCTACGAGAGTTA	300		
Db 241	AACGGAGCCCTTAANAAGAGAGCTGGCTGGTGCATTAATAATTATCATCTACGAGAGTTA	300		
Qy 301	ATGAATGATCCAGAAAGACACAGAGAAAGATTTCTTGCTCTTTGGAAACAATTCCTGAT	360		
Db 301	ATGAATGATCCAGAAAGACACAGAGAAAGATTTCTTGCTCTTTGGAAACAATTCCTGAT	360		
Qy 361	CGTTATAAGACATACCCGGAACCTCTATTTTGGAAATTCGATATCAACCTACGCGAAT	420		
Db 361	CGTTATAAGACATACCCGGAACCTCTATTTTGGAAATTCGATATCAACCTACGCGAAT	420		
Qy 421	CTTACTCCGGAATAATGGAATGAACTGCTTGAAGAGCTCTAAAGTTTAAAGTCAAT	480		
Db 421	CTTACTCCGGAATAATGGAATGAACTGCTTGAAGAGCTCTAAAGTTTAAAGTCAAT	480		
Qy 481	GACAAAAGCACAACATATATATATAGCACAAGCTGAATGGGGGGCTATATCTCCCTTGA	540		
Db 481	GACAAAAGCACAACATATATATATAGCACAAGCTGAATGGGGGGCTATATCTCCCTTGA	540		
Qy 541	AAACTGTCTGTCCCAAAATGGGAAAAAATCTATTAATTCACATTCACATCAATCTCT	600		
Db 541	AAACTGTCTGTCCCAAAATGGGAAAAAATCTATTAATTCACATTCACATCAATCTCT	600		
Qy 601	TTGGAATTTTACCATATCAGAGAGCTGAGGGGTGAAGAGCTAGAAATGGTTGGGAAGA	660		
Db 601	TTGGAATTTTACCATATCAGAGAGCTGAGGGGTGAAGAGCTAGAAATGGTTGGGAAGA	660		
Qy 661	AAGTGGGAGCTCCAGATGATCAGAAACATTTGTATAGAAATTCATTTTATAGAAGA	720		
Db 661	AAGTGGGAGCTCCAGATGATCAGAAACATTTGTATAGAAATTCATTTTATAGAAGA	720		
Qy 721	TGTCATAAANAAGACAAAGACCAATTTACATAGTGAAGTTGGTGCCTACAGAAAGCT	780		
Db 721	TGTCATAAANAAGACAAAGACCAATTTACATAGTGAAGTTGGTGCCTACAGAAAGCT	780		
Qy 781	GACCTGAATCAAGATTAATAATGAGACTCTTGTGCTGTCGCAAAATGAGAAAGAGAGA	840		
Db 781	GACCTGAATCAAGATTAATAATGAGACTCTTGTGCTGTCGCAAAATGAGAAAGAGAGA	840		
Qy 841	TGAGAGTGGGCACTACTGGGAATTTTGTTCGGTTTGTGTTTATGATCTGAGAAAA	900		
Db 841	TGAGAGTGGGCACTACTGGGAATTTTGTTCGGTTTGTGTTTATGATCTGAGAAAA	900		
Qy 901	ACCTGGAATAAAGATCTTTAGAGCTTTAATAGAGAGATATACATTGAATAA	954		


```

XX 28-DEC-1999.
PD
XX 16-OCT-1997; 97US-0951889.
PF
XX 23-AUG-1995; 95US-0518615.
PR
XX (DIVE-) DIVERSA CORP.
PA
XX Mathur EJ, Lam DE;
PI
XX WPI: 2000-105335/09.
DR
XX Endoglucanase with carboxymethyl cellulase activity from Thermotoga
PT matitima, useful for degrading biomass to fuels and chemicals
PS
XX Disclosure: Col 17; 14pp; English.
XX
XX Sequences AA246760-761 represents PCR primers for the amplification of
CC the cDNA encoding a T. matitima endoglucanase with carboxymethyl
CC cellulase activity. The enzyme CMCase is thermostable and used to
CC degrade carboxymethyl cellulose (CMC) or more generally any
CC oligosaccharide that contains beta-1,4-glycosidic bonds, particularly in
CC detergents, textiles, guar gum, animal feed, plant biomass or waste
CC products. The CMCase enzyme is especially used to convert plant biomass
CC to chemicals and fuels, and also for clarification and extraction of
CC juice (e.g. in the brewing industry). CMCase can also be used to raise
CC specific antibodies.
XX
XX Sequence 51 BP; 14 A; 6 C; 14 G; 17 T; 0 other;
SQ
Query Match 2.2%; Score 21; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATGGGTGTGATCCTTTGAA 21
DB 31 ATGGGTGTGATCCTTTGAA 51
RESULT 8
AAD15244
ID AAD15244 standard; DNA; 51 BP.
XX
XX AAD15244;
AC
XX 15-NOV-2001 (first entry)
DT
XX Thermotoga matitima CMCase DNA amplifying primer #1.
DE
XX
XX Carboxymethyl cellulase; CMCase; plant biomass; animal feed; detergent;
KW textile industry; waste treatment; brewing industry; immunogen;
KW fruit juice industry; PCR primer; ss.
XX
XX Thermotoga matitima.
OS
XX
XX Key Location/Qualifiers
FH misc_feature 28
FT /+tag= a
FT /note= "This base is absent in the sequence shown
FT as SEQ ID NO:3 in Column 14 of the specification"
XX
XX US6245547-B1.
PN
XX
XX 12-JUN-2001.
PD
XX
XX 27-DEC-1999; 99US-0472857.
PF
XX
XX 23-AUG-1995; 95US-0518615.
PR
XX 16-OCT-1997; 97US-0951889.
PR
XX (DIVE-) DIVERSA CORP.
PA
XX
XX

```

```

PI Mathur EJ, Lam DE;
XX
XX WPI: 2001-528380/58.
DR
XX Polynucleotide isolated from thermophile eubacteria useful for
PT degradation of cellulose encodes an enzyme having a fully defined amino
PT acid sequence
XX
XX Example 1; Column 17-18; 14pp; English.
PS
XX
XX The present sequence is a PCR primer used to amplify a DNA coding for
XX a thermostable enzyme, an endoglucanase having carboxymethyl cellulase
XX activity referred to as (CMCase).
CC
CC The CMCase enzyme is useful for the degradation of cellulose,
CC for conversion of plant biomass into fuels and chemicals, in the
CC textile industry, in detergents, in animal feed, in waste treatment,
CC in fruit juice/brewing industry for clarification and extraction of
CC juices, to generate probes for identifying sequences encoding similar
CC enzymes from other organisms, as immunogen to produce monoclonal or
CC polyclonal antibodies.
XX
XX Sequence 51 BP; 14 A; 6 C; 14 G; 17 T; 0 other;
SQ
Query Match 2.2%; Score 21; DB 22; Length 51;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATGGGTGTGATCCTTTGAA 21
DB 31 ATGGGTGTGATCCTTTGAA 51
RESULT 9
AAL35482
ID AAL35482 standard; cDNA; 530 BP.
XX
XX AAL35482;
AC
XX 08-JAN-2002 (first entry)
DT
XX
XX Human musculoskeletal system related polynucleotide SEQ ID NO 824.
DE
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antilucer;
KW vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200155367-A1.
PN
XX
XX 02-AUG-2001;
PD
XX
XX 17-JAN-2001; 2001WO-US01338.
PE
XX
XX 31-JAN-2000; 2000US-0179065.
PR
XX 04-FEB-2000; 2000US-0180628.
PR
XX 24-FEB-2000; 2000US-0184664.
PR
XX 02-MAR-2000; 2000US-0186350.
PR
XX 16-MAR-2000; 2000US-0189874.
PR
XX 17-MAR-2000; 2000US-0190076.
PR
XX 18-APR-2000; 2000US-0198123.
PR
XX 19-MAY-2000; 2000US-0205153.
PR
XX 07-JUN-2000; 2000US-0209467.
PR
XX 28-JUN-2000; 2000US-0214886.
PR
XX 30-JUN-2000; 2000US-0215135.
PR
XX 07-JUL-2000; 2000US-0216647.
PR
XX 07-JUL-2000; 2000US-0216880.
PR
XX 11-JUL-2000; 2000US-0217487.
PR
XX 11-JUL-2000; 2000US-0217496.
PR
XX 14-JUL-2000; 2000US-0218290.
PR

```

PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.

PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PR XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI, 2001-451937/48.
 XX P-PSDB; ABB03900.
 DR
 XX
 PT Isolated polypeptide for treating, preventing and/ or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 XX
 XX Claim 1; SEQ ID NO 824; 781pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (AAU34669-AAU37666) and proteins
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (anti)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,

CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pctl_sequences.
XX

SO Sequence 530 BP; 163 A; 101 C; 97 G; 167 T; 2 other;

Query Match 2.1%; Score 20; DB 22; Length 530;

Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 891 TCTGAGAAACCTGGAATA 910
DB 63 TCTGAGAAACCTGGAATA 82

RESULT 10

AAQ32472/c
ID AAQ32472 standard; DNA: 849 BP.

AAQ32472:

26-APR-1993 (first entry)

HCV NS2-NS4 clone MX25.

Clone; polypeptide; NS2-NS4; Hepatitis C; virus; HCV; serum; HC;

transcriptase; cDNA; primer; allele; ss.

Hepatitis C virus.

EP518313-A.

16-DEC-1992.

11-JUN-1992; 92EP-0109812.

11-JUN-1991; 91JP-0139268.

12-JUL-1991; 91JP-0172794.

07-OCT-1991; 91JP-0287008.

16-DEC-1991; 91JP-0333229.

20-APR-1992; 92JP-0099957.

(MITU) MITSUBISHI KASEI CORP.

Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;

Teranishi Y;

WPI: 1992-417213/51.

P-PSDB: AAR29559.

New hepatitis C virus gene and its encoded protein - used for

diagnosing and vaccinating against hepatitis C virus infections

Disclosure: Page 117-18; 305pp; English.

The sequences given in AAQ32472-82 and AAQ32442 are various clones which

encode the NS2-NS4 regions of the Hepatitis C Virus (HCV) gene of

the invention. These sequences were isolated from the serum of a

patient suffering from hepatitis C (HC). The isolated RNA sequences

of the primer sequences given in AAQ32553-64. The sequences were

then amplified using primer pairs. The cDNA sequences isolated

represent different alleles of the same region of the HCV gene.

Query Match 2.1%; Score 20; DB 13; Length 849;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 311 CAGAGAACACAGGAAGA 330
DB 125 CAGAGAACACAGGAAGA 106

RESULT 11

AAQ32483/c
ID AAQ32483 standard; DNA: 849 BP.

AAQ32483:

26-APR-1993 (first entry)

HCV NS2-NS4 clone MX25-2.

Clone; polypeptide; NS2-NS4; Hepatitis C; virus; HCV; serum; HC;

transcriptase; cDNA; primer; allele; ss.

Hepatitis C virus.

EP518313-A.

16-DEC-1992.

11-JUN-1992; 92EP-0109812.

11-JUN-1991; 91JP-0139268.

12-JUL-1991; 91JP-0172794.

07-OCT-1991; 91JP-0287008.

16-DEC-1991; 91JP-0333229.

20-APR-1992; 92JP-0099957.

(MITU) MITSUBISHI KASEI CORP.

Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;

Teranishi Y;

WPI: 1992-417213/51.

P-PSDB: AAR29852.

New hepatitis C virus gene and its encoded protein - used for

diagnosing and vaccinating against hepatitis C virus infections

Disclosure: Page 147-49; 305pp; English.

The sequences given in AAQ32483-501 are various clones which were used

in the isolation of the NS2-NS4 regions of the Hepatitis C Virus

(HCV) gene of the invention (see also AAQ32442 and AAQ32472-82). These

sequences were isolated from the serum of a patient suffering from

hepatitis C (HC). The isolated RNA sequences were converted into

cDNA using transcriptase. In the presence of one of the primer

sequences given in AAQ32578-79. The sequences were then amplified

using primer pairs. The cDNA sequences isolated represent different

alleles of the same region of the HCV gene. Sequence comparisons of

these clones showed that it is possible for a patient to carry more

than one HCV strain at one time. See also AAQ32436.

Sequence 849 BP; 150 A; 250 C; 247 G; 202 T; 0 other;

Query Match 2.1%; Score 20; DB 13; Length 849;

Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 311 CAGAGAACACAGGAAGA 330
DB 125 CAGAGAACACAGGAAGA 106

Sequence 849 BP; 132 A; 227 C; 232 G; 188 T; 70 other;

```

RESULT 12
AA032482/c
ID AA032482 standard; DNA: 849 BP.
XX
XX
AC AA032482;
XX
XX
DT 26-APR-1993 (first entry)
XX
XX
DE HCV NS2-NS4 clone MX25-1.
XX
XX
KM Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
KM transcriptase; cDNA; primer; allele; ss.
XX
XX
OS Hepatitis C virus.
XX
XX
PN EP518313-A.
XX
XX
PD 16-DEC-1992.
XX
XX
PF 11-JUN-1992; 92EP-0109812.
XX
XX
PR 11-JUN-1991; 91JP-0139268.
PR 12-JUL-1991; 91JP-0172794.
PR 07-OCT-1991; 91JP-0287008.
PR 16-DEC-1991; 91JP-0332329.
PR 20-APR-1992; 92JP-0099957.
XX
XX
PA (MITU ) MITSUBISHI KASEI CORP.
XX
XX
PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
XX
XX
DR WPI; 1992-417213/51.
DR P-PSDB; AAR29851.
XX
XX
PT New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections
XX
XX
PS Disclosure; Page 146-47; 305pp; English.
XX
XX
CC The sequences given in AA032472-82 and AA032442 are various clones which
CC encode the NS2-NS4 regions of the Hepatitis C Virus (HCV) gene of
CC the invention. These sequences were isolated from the serum of a
CC patient suffering from hepatitis C (HC). The isolated RNA sequences
CC were converted into cDNA using transcriptase in the presence of one
CC of the primer sequences given in AA032553-64. The sequences were
CC then amplified using primer pairs. The cDNA sequences isolated
CC represent different alleles of the same region of the HCV gene.
CC Sequence comparisons of these clones showed that it is possible for a
CC patient to carry more than one HCV strain at one time. See also
CC AA032436.
XX
XX
SQ Sequence 849 BP; 146 A; 245 C; 249 G; 209 T; 0 other;

Query Match 2.1%; Score 20; DB 13; Length 849;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
DB 125 CAGAGAACACAGGAAGA 106

RESULT 13
AA032484/c
ID AA032484 standard; DNA: 849 BP.
XX
XX
AC AA032484;
XX
XX
DT 26-APR-1993 (first entry)
XX
XX
DE HCV NS2-NS4 clone MX25-3.
XX

```

```

KM Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
KM transcriptase; cDNA; primer; allele; ss.
XX
XX
XX Hepatitis C virus.
XX
XX
PN EP518313-A.
XX
XX
PD 16-DEC-1992.
XX
XX
PF 11-JUN-1992; 92EP-0109812.
XX
XX
PR 11-JUN-1991; 91JP-0139268.
PR 12-JUL-1991; 91JP-0172794.
PR 07-OCT-1991; 91JP-0287008.
PR 16-DEC-1991; 91JP-0332329.
PR 20-APR-1992; 92JP-0099957.
XX
XX
PA (MITU ) MITSUBISHI KASEI CORP.
XX
XX
PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
XX
XX
DR WPI; 1992-417213/51.
DR P-PSDB; AAR29853.
XX
XX
PT New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections
XX
XX
PS Disclosure; Page 149-51; 305pp; English.
XX
XX
CC The sequences given in AA032483-501 are various clones which were used
CC in the isolation of the NS2-NS4 regions of the Hepatitis C Virus
CC (HCV) gene of the invention (see also AA032442 and AA032472-82). These
CC sequences were isolated from the serum of a patient suffering from
CC hepatitis C (HC). The isolated RNA sequences were converted into
CC cDNA using transcriptase in the presence of one of the primer
CC sequences given in AA032578-79. The sequences were then amplified
CC using primer pairs. The cDNA sequences isolated represent different
CC alleles of the same region of the HCV gene. Sequence comparisons of
CC these clones showed that it is possible for a patient to carry more
CC than one HCV strain at one time. See also AA032436.
XX
XX
SQ Sequence 849 BP; 146 A; 246 C; 247 G; 210 T; 0 other;

Query Match 2.1%; Score 20; DB 13; Length 849;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
DB 125 CAGAGAACACAGGAAGA 106

RESULT 14
AA032479/c
ID AA032479 standard; DNA: 1280 BP.
XX
XX
AC AA032479;
XX
XX
DT 26-APR-1993 (first entry)
XX
XX
DE HCV NS2-NS4 clone MX25026.
XX
XX
KM Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
KM transcriptase; cDNA; primer; allele; ss.
XX
XX
XX Hepatitis C virus.
XX
XX
PN EP518313-A.
XX
XX
PD 16-DEC-1992.
XX
XX
PF 11-JUN-1992; 92EP-0109812.
XX

```

```

XX 11-JUN-1991; 91JP-0139268.
PR 12-JUL-1991; 91JP-0172794.
PR 07-OCT-1991; 91JP-0287008.
PR 16-DEC-1991; 91JP-0332329.
PR 20-APR-1992; 92JP-0099957.
XX
PA (MITU ) MITSUBISHI KASEI CORP.
XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
XX
XX WPI: 1992-417213/51.
DR P-PSDB: AAR29848.
XX
XX New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections
XX
XX Disclosure: Page 129-31; 305pp; English.
XX
XX The sequences given in AAQ32472-82 and AAQ32442 are various clones which
CC encode the NS2-NS4 regions of the Hepatitis C Virus (HCV) gene of
CC the invention. These sequences were isolated from the serum of a
CC patient suffering from hepatitis C (HC). The isolated RNA sequences
CC were converted into cDNA using transcriptase in the presence of one
CC of the primer sequences given in AAQ3253-64. The sequences were
CC then amplified using primer pairs. The cDNA sequences isolated
CC represent different alleles of the same region of the HCV gene.
CC Sequence comparisons of these clones showed that it is possible for a
CC patient to carry more than one HCV strain at one time. See also
CC AAQ32436.
XX
SQ Sequence 1280 BP; 204 A; 353 C; 357 G; 270 T; 96 other;

Query Match 2.1%; Score 20; DB 13; Length 1280;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 311 CAGAAGACACAGAAGAAGA 330
DB 125 CAGAAGACACAGAAGAAGA 106

RESULT 15
AAQ32494/C
ID AAQ32494 standard; DNA; 1280 BP.
XX
XX AAQ32494;
XX
XX 26-APR-1993 (first entry)
XX
XX HCV NS2-NS4 clone MX25026A-1.
XX
XX Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
XX transcriptase; cDNA; primer; allele; ss.
XX
XX Hepatitis C virus.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1278
FT CDS /*tag= a
XX
XX EP518313-A.
XX
XX 16-DEC-1992.
XX
XX 11-JUN-1992; 92EP-0109812.
XX
XX 11-JUN-1991; 91JP-0139268.
XX 12-JUL-1991; 91JP-0172794.
XX 07-OCT-1991; 91JP-0287008.
XX 16-DEC-1991; 91JP-0332329.
XX 20-APR-1992; 92JP-0099957.

```

```

XX (MITU ) MITSUBISHI KASEI CORP.
XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
XX
XX WPI: 1992-417213/51.
DR P-PSDB: AAR29863.
XX
XX New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections
XX
XX Disclosure: Page 165-67; 305pp; English.
XX
XX The sequences given in AAQ32483-501 are various clones which were used
CC in the isolation of the NS2-NS4 regions of the Hepatitis C Virus
CC (HCV) gene of the invention (see also AAQ32442 and AAQ32472-82). These
CC sequences were isolated from the serum of a patient suffering from
CC hepatitis C (HC). The isolated RNA sequences were converted into
CC cDNA using transcriptase in the presence of one of the primer
CC sequences given in AAQ32578-79. The sequences were then amplified
CC using primer pairs. The cDNA sequences isolated represent different
CC alleles of the same region of the HCV gene. Sequence comparisons of
CC these clones showed that it is possible for a patient to carry more
CC than one HCV strain at one time. See also AAQ32436.
XX
SQ Sequence 1280 BP; 222 A; 385 C; 376 G; 297 T; 0 other;

Query Match 2.1%; Score 20; DB 13; Length 1280;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 311 CAGAAGACACAGAAGAAGA 330
DB 125 CAGAAGACACAGAAGAAGA 106

RESULT 16
AAQ32495/C
ID AAQ32495 standard; DNA; 1280 BP.
XX
XX AAQ32495;
XX
XX 26-APR-1993 (first entry)
XX
XX HCV NS2-NS4 clone MX25026B-1.
XX
XX Clone; polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
XX transcriptase; cDNA; primer; allele; ss.
XX
XX Hepatitis C virus.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1278
FT CDS /*tag= a
XX
XX EP518313-A.
XX
XX 16-DEC-1992.
XX
XX 11-JUN-1992; 92EP-0109812.
XX
XX 11-JUN-1991; 91JP-0139268.
XX 12-JUL-1991; 91JP-0172794.
XX 07-OCT-1991; 91JP-0287008.
XX 16-DEC-1991; 91JP-0332329.
XX 20-APR-1992; 92JP-0099957.
XX
XX (MITU ) MITSUBISHI KASEI CORP.
XX Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
XX

```


CC	healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC	and (f) infectious diseases such as viral, bacterial, fungal and
CC	parasitic infections.
CC	Note: The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at http://www.int/pub/published_pct_sequences .
CC	
SO	Sequence 1350 BP; 414 A; 250 C; 303 G; 383 T; 0 other;
Query Match	2.1%; Score 20; DB 22; Length 1350;
Best Local Similarity	100.0%; Pred. No. 10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
Oy	891 TCTGAGAAAACCTCGAATA 910
Db	866 TCTGAGAAAACCTCGAATA 885
RESULT 18	
AL37284	
ID	AL37284 standard; DNA; 1350 BP.
XX	AL37284;
AC	
XX	
DT	08-JAN-2002 (first entry)
XX	
DE	Human musculoskeletal system related polynucleotide SEQ ID NO 3649.
XX	
KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antilacer;
KW	vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW	cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW	neurological disease; infection; human; secreted protein;
KW	musculoskeletal system; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200155367-A1.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01338.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 08-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-451937/48.
 Isolated polypeptide for treating, preventing and/or prognosing
 disorders related to the musculoskeletal system including
 musculoskeletal cancers and also for testing and detection e.g.
 diagnosis -
 XX
 PS Example 2; SEQ ID NO 3649; 781pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune hemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 CC
 XX Sequence 1350 BP; 414 A; 250 C; 303 G; 383 T; 0 other;
 Query Match 2.1%; Score 20; DB 22; Length 1350;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 891 TCTGAGAAAACCTGATA 910
| | | | | | | | | | | | | | | | | |
Db 866 TCTGAGAAAACCTGATA 885

RESULT 19
AAL37285
ID AAL37285 standard; DNA: 1350 BP.
XX
AC AAL37285;
XX
DT 08-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 3650.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antileuc;
KW valnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.
XX
OS Homo sapiens.
XX
PN WO200155367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01338.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0218290.
PR 14-JUL-2000; 2000US-0218964.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225275.
PR 14-AUG-2000; 2000US-0225277.
PR 14-AUG-2000; 2000US-0225278.
PR 14-AUG-2000; 2000US-0225279.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236371.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239937.
PR 13-OCT-2000; 2000US-0239938.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241825.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 05-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256179.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-451937/48.
 XX
 PT Isolated polypeptide for treating, preventing and/or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 XX
 XX
 XX Example 2: SEQ ID NO 3650; 781bp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (AA034669-AA037666) and proteins
 CC (AB03087-AB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 1350 BP; 413 A; 250 C; 303 G; 384 T; 0 other;
 SQ
 Query Match 2.1%; Score 20; DB 22; Length 1350;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 891 TCTGAGAAAACCTGGAATA 910
 DB 866 TCTGAGAAAACCTGGAATA 885

RESULT 20
 AA032442/c
 ID AA032442 standard; DNA; 3564 BP.
 XX
 AC AA032442;
 XX
 DT 26-APR-1993 (first entry)
 XX

DE HCV NS2-NS4 clone MX25N15.
 XX
 KW Clone: polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
 KW transcriptase; cDNA; primer; allele; ss.
 OS Hepatitis C virus.
 OS
 PN EP518313-A.
 XX
 PD 16-DEC-1992.
 XX
 PF 11-JUN-1992; 92EP-0109812.
 XX
 PR 11-JUN-1991; 91JP-0139268.
 PR 12-JUL-1991; 91JP-0172794.
 PR 07-OCT-1991; 91JP-0287008.
 PR 16-DEC-1991; 91JP-0332329.
 PR 20-APR-1992; 92JP-0099957.
 XX
 PA (MITU) MITSUBISHI KASEI CORP.
 XX
 PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
 PI Teranishi Y;
 XX
 DR WPI: 1992-417213/51.
 DR P-PSDB: AAR29660.
 XX
 PT New hepatitis C virus gene and its encoded protein - used for
 PT diagnosing and vaccinating against hepatitis C virus infections
 PT
 PS Disclosure; Page 139-46; 305bp; English.
 XX
 CC The sequences given in AA032472-82 and AA032442 are various clones which
 CC encode the NS2-NS4 regions of the Hepatitis C virus (HCV) gene of
 CC the invention. These sequences were isolated from the serum of a
 CC patient suffering from Hepatitis C (HC). The isolated RNA sequences
 CC were converted into cDNA using transcriptase in the presence of one
 CC of the primer sequences given in AA032553-64. The sequences were
 CC then amplified using primer pairs. The cDNA sequences isolated
 CC represent different alleles of the same region of the HCV gene.
 CC Sequence comparisons of these clones showed that it is possible for a
 CC patient to carry more than one HCV strain at one time. See also
 CC AA032436.
 CC
 XX Sequence 3564 BP; 640 A; 1019 C; 1026 G; 747 T; 132 other;
 SQ

Query Match 2.1%; Score 20; DB 13; Length 3564;
 Best Local Similarity 100.0%; Pred. No. 9.8;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 311 CAGACAGACACAGGAAGA 330
 DB 125 CAGACAGACACAGGAAGA 106

RESULT 21
 AA032501/c
 ID AA032501 standard; DNA; 3564 BP.
 XX
 AC AA032501;
 XX
 DT 26-APR-1993 (first entry)
 XX
 DE HCV NS2-NS4 clone N25N15-1.
 XX
 KW Clone: polypeptide; NS2-NS4; Hepatitis C; Virus; HCV; serum; HC;
 KW transcriptase; cDNA; primer; allele; ss.
 OS Hepatitis C virus.
 OS
 PN EP518313-A.
 PD 16-DEC-1992.
 XX

```

XX 11-JUN-1992; 92EP-0109812.
PF 11-JUN-1991; 91JP-0139268.
PR 12-JUL-1991; 91JP-0172794.
PR 07-OCT-1991; 91JP-0287008.
PR 16-DEC-1991; 91JP-0332329.
PR 20-APR-1992; 92JP-0099957.
XX
PA (MITU ) MITSUBISHI KASEI CORP.
PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
DR WPI: 1992-417213/51.
DR P-PSDB; AAR29870.
XX
PT New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections
PS
XX Disclosure; Page 186-92; 305pp; English.
XX
CC The sequences given in AAQ32483-501 are various clones which were used
CC in the isolation of the NS2-NS4 regions of the Hepatitis C Virus
CC (HCV) gene of the invention (see also AAQ32442 and AAQ32472-82). These
CC sequences were isolated from the serum of a patient suffering from
CC hepatitis C (HC). The isolated RNA sequences were converted into
CC cDNA using transcriptase in the presence of one of the primer
CC sequences given in AAQ32578-79. The sequences were then amplified
CC using primer pairs. The cDNA sequences isolated represent different
CC alleles of the same region of the HCV gene. Sequence comparisons of
CC these clones showed that it is possible for a patient to carry more
CC than one HCV strain at one time. See also AAQ32436.
XX
SQ Sequence 3564 BP; 668 A; 1064 C; 1046 G; 786 T; 0 other;
Query Match 2.1%; Score 20; DB 13; Length 3564;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 CAGAGAACACACAGAGAGA 330
Db 125 CAGAGAACACACAGAGAGA 106
RESULT 22
ABL28818/C
ID ABL28818 standard; DNA; 4235 BP.
XX
AC ABL28818;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37927.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
PF 23-MAR-2001; 2001WO-US09231.
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PT Venter JC, Adams M, Li PWD, Myers EW;

```

```

DR WPI: 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 37927; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (ABB57737-ABB72072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pft_sequences.
XX
SQ Sequence 4235 BP; 1103 A; 850 C; 759 G; 1523 T; 0 other;
Query Match 2.1%; Score 20; DB 23; Length 4235;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 169 TCCCTCTATTAATCATGCG 208
Db 3940 TCCCTCTATTAATCATGCG 3921
RESULT 23
AAQ32436/C
ID AAQ32436 standard; DNA; 7911 BP.
XX
AC AAQ32436;
XX
DT 26-APR-1993 (first entry)
XX
DE HCV antigen clone T7N1-30.
XX
KW Clone; Hepatitis C Virus; HCV; core-envelope; NS1(gp70); NS2-NS4;
KW NS4-NS5; region; diagnostic method; antibody; suppress; control;
KW proteolytic; process; precursor; polypeptide; ss.
XX
OS Hepatitis C virus.
XX
FH
FT Key Location/Qualifiers
FT CDS 369..7898
FT /tag= a
XX
PN EP518313-A.
XX
PD 16-DEC-1992.
XX
PE 11-JUN-1992; 92EP-0109812.
XX
PR 11-JUN-1991; 91JP-0139268.
PR 12-JUL-1991; 91JP-0172794.
PR 07-OCT-1991; 91JP-0287008.
PR 16-DEC-1991; 91JP-0332329.
PR 20-APR-1992; 92JP-0099957.
XX
PA (MITU ) MITSUBISHI KASEI CORP.
XX
PI Hayashi N, Honda Y, Murakami T, Seki M, Takahashi K;
PI Teranishi Y;
DR WPI: 1992-417213/51.
DR P-PSDB; AAR29527.
XX
PT New hepatitis C virus gene and its encoded protein - used for
PT diagnosing and vaccinating against hepatitis C virus infections

```

XX Claim 1 and 3; Page 259-272; 305pp; English.
PS
XX
CC The sequences given in AAQ32436-539 are various clones which were all
CC derived from the isolated Hepatitis C Virus (HCV) gene of the
CC invention. AAQ32436 contains the entire protein gene, whereas AAQ32444-
CC 53 and AAQ32438-39, AAQ32454-68, AAQ32472-82 and AAQ32442, and
CC AAQ32502-12 and AAQ32443 encode the core-envelope, NS1(gp70), NS2-NS4
CC and NS4-NS5 regions respectively. AAQ32436 represents the entire gene
CC sequence. The HCV gene is useful in the development of a diagnostic
CC method which is more accurate and effective than conventional ones, in
CC the detection of antibodies raised against a wide range of HCVs which
CC have been hardly detected before. The complete gene may be used in an
CC in vitro screening system for a substance capable of specifically
CC suppressing or controlling a proteolytic processing of a precursor
CC polypeptide of HCV.
XX
SQ Sequence 7911 BP; 1556 A; 2363 C; 2292 G; 1700 T; 0 other;
Query Match 2.1%; Score 20; DB 13; Length 7911;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 CAGAGACACACAGGAAGA 330
|||
Db 2686 CAGAGACACACAGGAAGA 2667

Search completed: April 26, 2003, 11:33:26
Job time : 211 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 10:58:07 ; Search time 1108 Seconds

(without alignments)
13944.489 Million cell updates/sec

Title: US-09-880-729a-1

Perfect score: 954

Sequence: 1 atgggtgtgtatccttcta.....gagagatagcatgaataa 954

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 20

Total number of hits satisfying chosen parameters: 18

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST:*
1: em_estba:*
2: em_estlum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	2.4	563	10	AM308797 sft72a1.y
2	21	2.2	252	10	AV359581 AV359581
3	20	2.1	139	9	AV097171 AV097171
4	20	2.1	258	10	BB153117 BB153117
5	20	2.1	307	9	AA312956 EST183735
6	20	2.1	385	9	AI592815 mt70b05.y

7	20	2.1	486	17	A0653914
8 <td>20 <td>2.1 <td>509 <td>12 <td>BC081375</td> </td></td></td></td>	20 <td>2.1 <td>509 <td>12 <td>BC081375</td> </td></td></td>	2.1 <td>509 <td>12 <td>BC081375</td> </td></td>	509 <td>12 <td>BC081375</td> </td>	12 <td>BC081375</td>	BC081375
9 <td>20 <td>2.1 <td>608 <td>10 <td>AW963311</td> </td></td></td></td>	20 <td>2.1 <td>608 <td>10 <td>AW963311</td> </td></td></td>	2.1 <td>608 <td>10 <td>AW963311</td> </td></td>	608 <td>10 <td>AW963311</td> </td>	10 <td>AW963311</td>	AW963311
10 <td>20 <td>2.1 <td>621 <td>10 <td>AV386359</td> </td></td></td></td>	20 <td>2.1 <td>621 <td>10 <td>AV386359</td> </td></td></td>	2.1 <td>621 <td>10 <td>AV386359</td> </td></td>	621 <td>10 <td>AV386359</td> </td>	10 <td>AV386359</td>	AV386359
11 <td>20 <td>2.1 <td>673 <td>17 <td>A0577386</td> </td></td></td></td>	20 <td>2.1 <td>673 <td>17 <td>A0577386</td> </td></td></td>	2.1 <td>673 <td>17 <td>A0577386</td> </td></td>	673 <td>17 <td>A0577386</td> </td>	17 <td>A0577386</td>	A0577386
12 <td>20 <td>2.1 <td>713 <td>12 <td>BC068343</td> </td></td></td></td>	20 <td>2.1 <td>713 <td>12 <td>BC068343</td> </td></td></td>	2.1 <td>713 <td>12 <td>BC068343</td> </td></td>	713 <td>12 <td>BC068343</td> </td>	12 <td>BC068343</td>	BC068343
13 <td>20 <td>2.1 <td>769 <td>17 <td>A0914192</td> </td></td></td></td>	20 <td>2.1 <td>769 <td>17 <td>A0914192</td> </td></td></td>	2.1 <td>769 <td>17 <td>A0914192</td> </td></td>	769 <td>17 <td>A0914192</td> </td>	17 <td>A0914192</td>	A0914192
14 <td>20 <td>2.1 <td>771 <td>17 <td>AC0871693</td> </td></td></td></td>	20 <td>2.1 <td>771 <td>17 <td>AC0871693</td> </td></td></td>	2.1 <td>771 <td>17 <td>AC0871693</td> </td></td>	771 <td>17 <td>AC0871693</td> </td>	17 <td>AC0871693</td>	AC0871693
15 <td>20 <td>2.1 <td>784 <td>17 <td>CNS0220R</td> </td></td></td></td>	20 <td>2.1 <td>784 <td>17 <td>CNS0220R</td> </td></td></td>	2.1 <td>784 <td>17 <td>CNS0220R</td> </td></td>	784 <td>17 <td>CNS0220R</td> </td>	17 <td>CNS0220R</td>	CNS0220R
16 <td>20 <td>2.1 <td>851 <td>17 <td>A2749869</td> </td></td></td></td>	20 <td>2.1 <td>851 <td>17 <td>A2749869</td> </td></td></td>	2.1 <td>851 <td>17 <td>A2749869</td> </td></td>	851 <td>17 <td>A2749869</td> </td>	17 <td>A2749869</td>	A2749869
17 <td>20 <td>2.1 <td>873 <td>17 <td>CNS045PV</td> </td></td></td></td>	20 <td>2.1 <td>873 <td>17 <td>CNS045PV</td> </td></td></td>	2.1 <td>873 <td>17 <td>CNS045PV</td> </td></td>	873 <td>17 <td>CNS045PV</td> </td>	17 <td>CNS045PV</td>	CNS045PV
18 <td>20 <td>2.1 <td>1539 <td>11 <td>AK012307</td> </td></td></td></td>	20 <td>2.1 <td>1539 <td>11 <td>AK012307</td> </td></td></td>	2.1 <td>1539 <td>11 <td>AK012307</td> </td></td>	1539 <td>11 <td>AK012307</td> </td>	11 <td>AK012307</td>	AK012307

ALIGNMENTS

RESULT 1
AM308797
LOCUS
DEFINITION
sft72a1.y1 Gm-cl013 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl013-5109.5, similar to SW:CAHC_PEA P17067 CARBONIC ANHYDRASE, CHLOROPLAST PRECURSOR ;, mRNA sequence.
ACCESSION
AM308797
VERSION
AM308797.1 GI:6724398
KEYWORDS
EST.
SOURCE
soybean.
ORGANISM
Glycine max

REFERENCE
AUTHORS
Shoemaker,R., Kelm,P., Vodkin,L., Erpeliding,J., Corryell,V., Khanna,A., Bolla,B., Matra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Rilter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Willson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estelw@wustl.edu

TITLE
JOURNAL
COMMENT
This clone is available through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccuteresgen.com
Seq primer: -40RP from cldco
High quality sequence stop: 389.
Location/Qualifiers
1..563
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl013-5109"
/clone_11b="Gm-cl013"
/tissue_type="Whole seedlings, 2-3 week old seedlings, greenhouse grown"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site: 1: EcoRI; Site: 2: XhoI. This cDNA library was constructed from mRNA isolated from whole seedlings of 2-3 week old greenhouse grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dI) sequence with a blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments


```

/clone="2410026A20"
/clone_lib="Mus musculus C57BL/6J ES cell"
/cell_type="ES cell"
/note="organ: mammary gland; Vector: p773D-Pac (Pharmacia)
with a modified polylinker. Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5']
TGTTCACATCTGAGTGGAGCGCGCGCATGTTTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT      24 a      26 c      34 g      55 t
ORIGIN
Query Match      2.1%; Score 20; DB 9; Length 139;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 726 AAAAAAGACAAAGACCA 745
Db 100 AAAAAAGACAAAGACCA 81

RESULT 4
BBI53117/c      258 bp      mRNA      linear      EST 28-JUN-2000
BBI53117 RIKEN full-length enriched, 16 days neonate thymus Mus
musculus cDNA clone A130010K07 3' similar to U88328 Mus musculus
suppressor of cytokine signalling-3 (SOCS-3) mRNA, mRNA sequence.
BBI53117
BBI53117.1 GI:8808054
EST
house mouse.
Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 258)
Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci,
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hizane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koyas, S., Kurihara, C., Kusabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya,
T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Kono, H., et al.)
Unpublished (2000)
Contact: Yoshinori Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL: http://genome-gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagakawa, S., Sasaki,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermolabile and thermocatalytic of thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)

```

```

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source
Location/Qualifiers
1..258
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A130010K07"
/clone_lib="RIKEN full-length enriched, 16 days neonate
thymus"
/tissue_type="thymus"
/dev_stage="16 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5']
GAGAGAGAGAGATCCAGAGCGCTCTTTTCTTTTCTTTTCTTTTCTTTT 3', cDNA was
transcribed by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adaptor of
sequence [5'] GAGAGAGAGATTCGATTAATTAATTAATTCCTCCCTCC
3'. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from lambda
FLC I."
BASE COUNT      77 a      48 c      37 g      95 t      1 others
ORIGIN
Query Match      2.1%; Score 20; DB 10; Length 258;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 TATTTTGAATTCGAAT 405
Db 188 TATTTTGAATTCGAAT 169.

RESULT 5
AA312956      307 bp      mRNA      linear      EST 19-APR-1997
DEFINITION
AA312956 Bone VII Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION
AA312956.1 GI:1965304
VERSION
AA312956.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 307)
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult,
C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White,
O., Sutton, G., Blake, J.A., Brannon, R.C., Man-We, C., Clayton, R.A.,
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald,
L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S., Glodak, A.,
Guelm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palanges, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H., Raymond, L.,
Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon,
M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
Venter, J.C.

```

TITLE	Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL	Nature 377 (6547 Suppl.), 3-174 (1995)
MEDLINE	96026280
COMMENT	Other ESTs: THC183992
Contact:	Kerlavange, AR

initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl.), 3-174 (1995)
96026280 Other_ESTs: TNC183992
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavet@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tcdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers

FEATURES
source

1. :307

```
/db_xref="ATCC (1host):167970"
```

```
/clone_lib="Bone VII"
```

```

/note="Organ: bone; Vector: pBluescript SK-; Site_1: EcoRI

```

BASE COUNT	98 a	58 c	52 g	95 t	4 others
ORIGIN					

Query Match	2.1%	Score 20	DB 9	Length 307
Best Local Similarity	100.0%	Pred. No. 27		
Matches	20	Conservative	0	Mismatches 0
				Indels 0
				Gaps 0

QY 891 TCTGAGAAAAACCTGGAATA 910
 |||||
 Db 111 TCTGAGAAAAACCTGGAATA 130

RESULT 6				
AI592815/c	385 bp	mRNA	linear	EST 15-MAR-2000
LOCUS				
DEFINITION				
IMAGE:635217 5', mRNA sequence.				
AI592815				
mt70b05.y1 Soares mouse lymph node NBLMN Mus musculus cDNA clone				

VERSION	AI592815.1	GI:4601863
KEYWORDS	EST.	
SOURCE	house.mouse.	

REFERENCE
AUTHORS

1 (bases 1 to 385)
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., ... et al.

TITLE	The Washu-NCI Mouse EST Project 1999
JOURNAL	Unpublished (1999)
COMMENT	Contact: Marra M/Washu-NCI Mouse EST Project 1999

Email: mouseestewartson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.lnl.gov) for further information. This read is a RESEQUENCE of a previously sequenced mouse clone. This read has been verified (found to hit its original self in the correct orientation)
Putative full length read
vector to vector length is 391
MGI:387209
Seq primer: -40RP from Gbbco
high quality sequence stop: 371

POLYA=No.	Location/Qualifiers
FEATURES	1. .385
source	/organism="Mus musc

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1b="IMAGE:655217"
/clone_1b="Soares lymph node NBLMLN"
/sex="male"
/tissue_type="lymph node"

```

```
/dev_stage-4 weeks
/lab host="DH10B"
```

```

/notes=Ufgall: 1ymphn code; vector: p113b fac (fthlmaccn)
with a modified polylinker; site_1: Not I; site_2: Eco RI;

```

1st strand cdna was primed with a Not I - oligo(a1) primer

3'1': double-stranded CPNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA

provided by Dr. Bertrand Jordan. Library constructed and normalized by Bento Soares and M. Fatima Bonaldó."

a	79 c	76 g	131 t
---	------	------	-------

2 19: Eucoro 30: DB 9: Ierath 385:

Query Match	2.1%;	Score 20;	DB 9;	Length 385;
Best Local Similarity	100.0%;	Pred. No. 28;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY 726 AAAAAGAACCAAAAGACC 745
 |||||
 Db 46 AAAAAGAACCAAAAGACC 27

RESULT 7	486 bp	DNA	linear	655 22-JUN-1995
AO653914				
LOCUS				
DEFINITION	AO653914	486 bp	DNA	linear
	Sheared DNA-8D15	TR	Sheared DNA	Trypanosoma brucei genomic clone
	Sheared DNA-8D15,			DNA sequence.

KEYWORDS	GSS.
SOURCE	Trypanosoma brucei.
ORGANISM	Trypanosoma brucei

REFERENCE
AUTHORS
1 (bases 1 to 486)
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J.

TITLE	Frisel, C. and Adams, P. Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library
JOURNAL	Unpublished (1999)
COMMENT	Other-GSS: Sheared DNA-BD15.TF 19990101

Department Of Biotechnology Genomics
The Institute For Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: nelsayed@tigr.org
Clones are derived from the *Trypanosoma brucei* Gurat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ARCC. Shared DNA end sequences search page:
<http://www.tigr.org/rdm/db/tbbdb/>.
Seq primer: M13-Reverse
Class: Sholgun.

FEATURES	Location/Qualifiers
source	1. .486

```
/organism="Trypanosoma brucei"  
/strain="TREU927/4 GUTat 10.1"  
/db_xref="taxon:5691"  
/clone="Sheared DNA-8D15"
```

```

/clone_lib="Sheared DNA"
/Note="Vector: pUC18; Site_1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 CUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + 1 method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Barrell, Oxford University
Press, 1999)."
BASE COUNT      93 a      115 c      128 g      150 t
ORIGIN
Query Match      2.1%; Score 20; DB 17; Length 486;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      681 TCAGAAACATTGTAGTAGAG 700
|||||
Db      391 TCAGAAACATTGTAGTAGAG 410

RESULT 8
BG081375      509 bp      mRNA      linear      EST 26-JAN-2001
LOCUS      H3064D12-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
DEFINITION      H3064D12-5', mRNA sequence.
ACCESSION      BG081375
VERSION      BG081375.1 GI:12563943
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 509)
Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka
, T.S., Carter, M.G. and Ko, M.S.H.
Verification and initial annotation of NIA mouse 15K cDNA clone set
unpublished (2001)
Other_FEATURES: H3064D12-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdnagisun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3064 row: D column: 12
Seq primer: -21M13 Reverse
High quality sequence stop: 509
POLYA-No.
FEATURES
source
Location/Qualifiers
1..509
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="niaEST:H3064D12-5"
/db_xref="taxon:10090"
/clone="H3064D12"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA
libraries"
/lab_host="DH10B"
/Note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearranged set of 15,247 clones from 11
embryo cDNA libraries (including preimplantation stage
embryos from unfertilized egg to blastocyst, embryonic
part of E7.5 embryos, extraembryonic part of E7.5 embryos
, and E12.5 female mesonephros/gonad) and one newborn
ovary cDNA library. Average insert size 1.5 kb. All
source libraries are cloned unidirectionally with Oligo(dT

```

```

)-Not primers. References include: (1) Genome-wide
expression profiling of mid-gestation placenta and embryo
using a 15,000 mouse developmental cDNA microarray, 2000,
Proc. Natl. Acad. Sci. U S A. 97: 9127-9132; (2)
Large-scale cDNA analysis reveals phased gene expression
patterns during preimplantation mouse development, 2000,
Development, 127: 1737-1749; (3) Genome-wide mapping of
unselected transcripts from extraembryonic tissue of
7.5-day mouse embryos reveals enrichment in the t-complex
and under-representation on the X chromosome, 1998, Hum
Mol Genet 7: 1967-1978."
BASE COUNT      176 a      81 c      90 g      160 t      2 others
ORIGIN
Query Match      2.1%; Score 20; DB 12; Length 509;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      336 TGCTCTTGGAAACAATG 355
|||||
Db      289 TGCTCTTGGAAACAATG 308

RESULT 9
AW963311      608 bp      mRNA      linear      EST 01-JUN-2000
LOCUS      AW963311
DEFINITION      EST375384 MAGE resequences, MACH Homo sapiens cDNA, mRNA sequence.
ACCESSION      AW963311
VERSION      AW963311.1 GI:8153147
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 608)
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt
, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
Quackenbush, J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnd@tigr.org
Plate: 186
Seq primer: Reverse.
FEATURES
source
Location/Qualifiers
1..608
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MACH"
/Note="Vector: pBluescriptSKm"
BASE COUNT      191 a      114 c      117 g      186 t
ORIGIN
Query Match      2.1%; Score 20; DB 10; Length 608;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      891 TCAGAAACCGGAGATA 910
|||||
Db      111 TCAGAAACCGGAGATA 130

RESULT 10
AV386359/c      621 bp      mRNA      linear      EST 27-OCT-1999
LOCUS      AV386359
DEFINITION      AV386359 Halocynthia roretzi Petillized egg Halocynthia roretzi
cDNA clone 009604_3 3', mRNA sequence.
ACCESSION      AV386359

```

```

VERSION      AV386359.1  GI:6131416
KEYWORDS     EST.
SOURCE       Halocynthia roretzi.
ORGANISM     Halocynthia roretzi
              Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
              Stolidobranchia; Pyuridae; Halocynthia.
REFERENCE    1 (bases 1 to 621)
AUTHORS      Makabe,K.W.
TITLE        Halocynthia roretzi EST
JOURNAL      Unpublished (1999)
COMMENT      Contact: Kazuhiko W. Makabe
              Department of Zoology, Graduate School of Science
              Kyoto University
              Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
              Tel: 81-75-753-4095
              Fax: 81-75-705-1113
              Email: kwmakabe@sci.kyoto-u.ac.jp.

FEATURES
SOURCE       1. 621
              /organism="Halocynthia roretzi"
              /db_xref="taxon:7729"
              /clone="009G04_3"
              /clone_lib="Halocynthia roretzi Fertilized egg"
              /dev_stage="Fertilized egg"
              /note="Organ: embryo"

BASE COUNT   196 a 99 c 102 g 224 t

ORIGIN
Query Match 2.1%; Score 20; DB 10; Length 621;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 384 TCTATTTTGAATTCGA 403
      |||||||
Db 465 TCTATTTTGAATTCGA 446

RESULT 11
AOS77386 673 bp DNA linear GSS 02-JUN-1999
LOCUS     nbxb0090P08f CUGI Rice BAC Library Oryza sativa genomic clone
DEFINITION nbxb0090P08f, DNA sequence.
ACCESSION AOS77386
VERSION    AOS77386.1 GI:4977871
KEYWORDS   GSS.
SOURCE     Oryza sativa.
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Ehrhartoideae; Oryzaceae; Oryza.
              1 (bases 1 to 673)
REFERENCE  Wing,R.A. and Dean,R.A.
AUTHORS    A BAC End Sequencing Framework to Sequence the Rice Genome
TITLE      Unpublished (1998)
JOURNAL    Contact: Wing RA
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: twing@clemson.edu
            Seg primer: TAATACGACTCACTATAGG
            Class: BAC ends
            High quality sequence stop: 280.

FEATURES
SOURCE     Location/Qualifiers
            1. 673
              /organism="Oryza sativa"
              /strain="Japonica"
              /cultivar="Nipponbare"
              /db_xref="taxon:4530"
              /clone="nbxb0090P08f"
              /clone_lib="CUGI Rice BAC Library"
              /tissue_type="leaf"

```

```

/lab_host="E. coli DH10B"
/note="Vector: pBelorAC11; Site 1: HindIII; Site 2:
HindIII; Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9%. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."

BASE COUNT   231 a 121 c 152 g 169 t

ORIGIN
Query Match 2.1%; Score 20; DB 17; Length 673;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 TTTGAAAGGACAAATAT 34
      |||||||
Db 598 TTTGAAAGGACAAATAT 617

RESULT 12
BG068343/C 713 bp mRNA linear EST 26-JAN-2001
LOCUS     H3064D12-3 N1A Mouse 15K cDNA Clone set Mus musculus cDNA clone
DEFINITION H3064D12-3', mRNA sequence.
ACCESSION BG068343
VERSION    BG068343.1 GI:12550912
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
            T.S., Carter,M.G. and Ko,M.S.H.
            Verification and initial annotation of N1A mouse 15K cDNA clone set
            Unpublished (2001)
            Other ESTs: H3064D12-5
            Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cnae@sgs.nci.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://19sun.9rc.nia.nih.gov/cDNA/15K.html for details.
            Plate: H3064 row: D column: 12
            Seg primer: -21M13 Forward
            High quality sequence stop: 713
            POLYA=yes.

FEATURES
SOURCE     Location/Qualifiers
            1. 713
              /organism="Mus musculus"
              /strain="C57Bl/6J"
              /db_xref="nlaEST:H3064D12-3"
              /db_xref="taxon:10090"
              /clone="H3064D12"
              /clone_lib="N1A Mouse 15K cDNA Clone set"
              /sex="Clones arrayed from a variety of cDNA libraries"
              /dev_stage="Clones arrayed from a variety of cDNA
              libraries"
              /lab_host="DH10B"

```

/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos , and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT) -Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum. Mol. Genet 7: 1967-1978."

BASE COUNT 220 a 112 g 255 t 1 others
ORIGIN

Query Match 2.1%; Score 20; DB 12; Length 713;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 336 TGCCTTTGGAACAATG 355
|||||
Db 487 TGCCTTTGGAACAATG 468

RESULT 13
AO914192 769 bp DNA linear GSS 02-DEC-1999
LOCUS nbe004/H17f CUGI Rice BAC Library (ECORI) Oryza sativa genomic
DEFINITION clone nbe004/H17f, DNA sequence.
ACCESSION AO914192
VERSION AO914192.1 GI:6510708
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 769)
AUTHORS Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: TATACGACTCAGCTATAGCG
Class: BAC ends
High quality sequence start: 51
High quality sequence stop: 412.

FEATURES
source
1..769
Location/Qualifiers
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbe004/H17f"
/clone_lib="CUGI Rice BAC Library (ECORI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: PBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely

on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 242 a 153 c 161 g 212 t 1 others
ORIGIN

Query Match 2.1%; Score 20; DB 17; Length 769;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 450 TGAGGAGCTCTAAAGTTA 469
|||||
Db 277 TGAGGAGCTCTAAAGTTA 296

RESULT 14
AO871693 771 bp DNA linear GSS 03-NOV-1999
LOCUS nbe0044J03r CUGI Rice BAC Library (ECORI) Oryza sativa genomic
DEFINITION clone nbe0044J03r, DNA sequence.
ACCESSION AO871693
VERSION AO871693.1 GI:6222144
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 771)
AUTHORS Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: GGAAACGCTATGACCATG
Class: BAC ends
High quality sequence start: 28
High quality sequence stop: 416.

FEATURES
source
1..771
Location/Qualifiers
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbe0044J03r"
/clone_lib="CUGI Rice BAC Library (ECORI)"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: PBACindigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a

haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*. The Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 240 a 153 c 164 g 212 t 2 others

Query Match 2.1%; Score 20; DB 17; Length 771;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 450 TGAGGAGCTCAAAAGTTA 469
|||||
Db 309 TGAGGAGCTCAAAAGTTA 328

RESULT 15
CNS02ZOR/C 784 bp DNA linear GSS 15-MAY-2000
DEFINITION Tetradodon nigroviridis genome survey sequence T7 end of clone 183K19 of library 6 from Tetradodon nigroviridis, genomic survey sequence.

ACCESSION AL221292.1 GI:7880111
VERSION 1
KEYWORDS GSS: genome survey sequence.
SOURCE Tetradodon nigroviridis.
ORGANISM Tetradodon nigroviridis.

REFERENCE Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Sautin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using Tetradodon nigroviridis DNA sequence

JOURNAL Unpublished
TITLE 2 (bases 1 to 784)
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Sautin, W., Bernot, A. and Weissenbach, J.

JOURNAL Unpublished
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradodon nigroviridis
REFERENCE 3 (bases 1 to 784)
AUTHORS Genoscope.

JOURNAL Direct Submission
TITLE Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetradodon>.

FEATURES
source Location/Qualifiers
1..784

BASE COUNT 226 a 148 c 160 g 250 t
ORIGIN /organism="Tetradodon nigroviridis"
/db_xref="taxon:99883"
/clone_1ib="G"
/clone_1ib="G"
/note="Genoscope sequence ID : CONAG183AF10LPI-end : T7"

Query Match 2.1%; Score 20; DB 17; Length 784;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 549 TGTCCTCAAAATGGGAAAAA 568
|||||
Db 605 TGTCCTCAAAATGGGAAAAA 586

RESULT 16
LOCUS A2749869 851 bp DNA linear GSS 25-JAN-2001
DEFINITION RPCI-24-97H19.T7 RPCI-24 Mus musculus genomic clone RPCI-24-97H19, DNA sequence.

ACCESSION A2749869 GI:12535028
VERSION A2749869.1
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 851)
AUTHORS Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akirret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorjys, E., Russell, D., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-24

JOURNAL Unpublished (1999)
TITLE Other_GSSs: RPCI-24-97H19.TV
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end Plate: 97 row: H column: 19
Seq primer: SP6
Clas: BAC ends.

FEATURES
source Location/Qualifiers
1..851

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1ib="RPCI-24-97H19"
/clone_1ib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 307 a 143 c 103 g 298 t

Query Match 2.1%; Score 20; DB 17; Length 851;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 336 TGCTCTTTGGAAACAAATTG 355
|||||
Db 799 TGCTCTTTGGAAACAAATTG 818

RESULT 17
CNS04SPV/C 873 bp DNA linear GSS 24-MAY-2000
LOCUS CNS04SPV

DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 007P20 of library H from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL305500

VERSION AL305500.1 GI:8198207

KEYWORDS GSS; genome survey sequence.

SOURCE Tetraodon nigroviridis.

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 873)

Auhot Rost-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Sautin, W., and Weissenbach, J.

Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

TITLE Unpublished

JOURNAL 2 (bases 1 to 873)

REFERENCE Unpublished

AUTHORS Rost-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Sautin, W., Bernot, A. and Weissenbach, J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL 3 (bases 1 to 873)

REFERENCE Unpublished

AUTHORS Direct Submission

TITLE Submitted (12-Apr-2000)

COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source

1.873

/organism="Tetraodon nigroviridis"

/db_xref="taxon:99883"

/clone="007P20"

/clone.lib="H"

/note="Genoscope sequence ID : COBH007DH10XD1-end : T7"

BASE COUNT 204 a 232 c 197 g 239 t 1 others

ORIGIN

Query Match 2.1%; Score 20; DB 17; Length 873;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 GGAGACTGGGAGTGTGTAT 101

|||||

Db 662 GGAGACTGGGAGTGTGTAT 643

RESULT 18

AK012307/c 1539 bp mRNA linear HTC 19-JAN-2002

LOCUS AK012307/c

DEFINITION Mus musculus 11 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2700031606;trans-prenyltransferase, full insert sequence.

ACCESSION AK012307

VERSION AK012307.1 GI:12848964

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) 11 days embryo cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20493374

PUBMED 11042159

REFERENCE 3

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishina, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE 4

AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batilov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiya, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stabelli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bona, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Rato, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Wetz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409 (6821), 685-690 (2001)

MEDLINE 21085660

PUBMED 11217851

REFERENCE 5 (bases 1 to 1539)

AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akinura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shikata, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

TITLE Submitted (10-Jul-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGGAGAAGGAAGATCCCAAGACGTCTTTTTCCTTTTN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGATTCTCGACTATTATTAATATCACCCTCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

FEATURES

Source

Location/Qualifiers

1. 1539

```
/organism="Mus musculus"
```

```
/strain="C57BL/6J"
```

```
/db_xref="FANTOM_DB:2700031G06"
```

/db_xref="MGD:MGI:1896014"

```
/db_xref="taxon:10090"
```

```
/clone="2/00031G06"
/c)one lib="RTKEN full"
```

```
/dev stage="11 days embryo"
```

257. .880

CDS

257. .88

```
/note="data source:MGD, source key:MG1:18892/8,
uri:demo:ISS
```

evidence:155
mutative

trans-prenyltransferase^m
pulcative

```

/codon_start=1

```

```
/protein_id="B
```

```
/db_xref="GI:12848965"
```

/db_xref="MGD:MGI:188244" /taxon="MSCROT"

/LIANSLALION- MCCCKYIYBSEN
 MSEYXFDPGKGKAFRIIVVLMARACN

VIDDASSRGGHTVNKIMGEKKAVLAGDILLSAASVALARIGNTAVVSMLAQVIEDLV

RGEFLQ LGSKENENERFAHYLEKTFKKTASLIANSCKAVCTFCLSS"

1517, .1522

```
/note="puta
1E30
```

polyA_signal

1517. .1522

```

/note="putative"
1E39

```

1539
/not

328 C 317

ORIGIN

ORIGIN

Query Match	Score	DB	Length
2.18;	20;	11;	1539;

Best Local Similarity 100.08; Pred. No. 31,

Matches	20;	conservative	0;	mismatches	0;	inliers	0;	gaps	0;
---------	-----	--------------	----	------------	----	---------	----	------	----

QY	310	CCAGAGACACACGGAAG	325

Db 1030 CCAGAGAACACAGGAAG 1011

Search completed: April 26, 2003, 11:51:59
Job time : 1116 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 11:23:08 ; Search time 46 Seconds
(without alignments)
6360.209 Million cell updates/sec

Title: US-09-880-729A-1
Perfect score: 954
Sequence: 1 atgggtgtgaccccttga.....gagggagatagcatgataa 954

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 20

Total number of hits satisfying chosen parameters: 8

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/prodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/prodata/1/lna/CTUS.COMB.seq:*
6: /cgn2_6/prodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	954	100.0	954	2	US-09-066-075-1
2	954	100.0	954	2	US-08-518-615A-1
3	954	100.0	954	3	US-08-951-889-1
4	954	100.0	954	2	US-09-472-857-1
5	21	2.2	51	2	US-09-066-075-3
6	21	2.2	51	2	US-08-518-615A-3
7	21	2.2	51	3	US-08-951-889-3
8	21	2.2	51	4	US-09-472-857-3

ALIGNMENTS

RESULT 1
US-09-066-075-1
; Sequence 1, Application US/09066075
; Patent No. 5925749
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILLILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY

COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,075
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-09-066-075-1

Query Match 100.0%; Score 954; DB 2; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGGGTGTGATCCCTTTTGAAGCAAAATATTTGGGAAGAGCATTAATATAGAAAT	60
DB	1	ATGGGTGTGATCCCTTTTGAAGCAAAATATTTGGGAAGAGCATTAATATAGAAAT	60
QY	61	GGCGTTGAACACCAATATGAGGAGCTGGGAGTGTGATTAAGATCTTCTTGAC	120
DB	61	GGCGTTGAACACCAATATGAGGAGCTGGGAGTGTGATTAAGATCTTCTTGAC	120
QY	121	ATTATTAAGAGCCGCTTCTCTCATGTTGCAATTCATTAAGATGATAGCAGCT	180
DB	121	ATTATTAAGAGCCGCTTCTCTCATGTTGCAATTCATTAAGATGATAGCAGCT	180
QY	181	TACGCGTTCTCTTATTAATATCATGATCGCTTTCTCAAAAGAGTGATGAAGTATA	240
DB	181	TACGCGTTCTCTTATTAATATCATGATCGCTTTCTCAAAAGAGTGATGAAGTATA	240
QY	241	AAGGAGCCCTGAAAGAGAGAGCGCTGCTATTAATATTCATCTACGAGAGTTA	300
DB	241	AAGGAGCCCTGAAAGAGAGAGCGCTGCTATTAATATTCATCTACGAGAGTTA	300
QY	301	ATGAATGATCCAGAGAACAACAAGAGATTTCTGCTTTGGAACAATTTGCTGAT	360
DB	301	ATGAATGATCCAGAGAACAACAAGAGATTTCTGCTTTGGAACAATTTGCTGAT	360
QY	361	CGTTATTAAGACTATCCCGAAGCTATTTTGAATTTGAAATTCGATGAACTCAGGAAAT	420
DB	361	CGTTATTAAGACTATCCCGAAGCTATTTTGAATTTGAAATTCGATGAACTCAGGAAAT	420
QY	421	CTTACGCCGGAATAATGAAATGAACTGCTGAGAGCCCTTAAGATTAAGATCAATT	480
DB	421	CTTACGCCGGAATAATGAAATGAACTGCTGAGAGCCCTTAAGATTAAGATCAATT	480
QY	481	GACAAAAGCACACTAATATATAGGCACAGTGAATGGGGGTATATTCGCTTGAA	540
DB	481	GACAAAAGCACACTAATATATAGGCACAGTGAATGGGGGTATATTCGCTTGAA	540
QY	541	AAACTGTCTGCCCAAAATGGGAAATTTCTATAGTTACAAATTCACATCAATCCT	600
DB	541	AAACTGTCTGCCCAAAATGGGAAATTTCTATAGTTACAAATTCACATCAATCCT	600

Qy	601	TTCAATTACC	CCATCAAGAGCTG	AGTGGGTGAAGACATCTG	AGAAATGTTGGCAAG	660
Qy	601	TTCAATTACC	CCATCAAGAGCTG	AGTGGGTGAAGACATCTG	AGAAATGTTGGCAAG	660
Db	601	TTCGAATTAC	CCATCAAGAGCTG	AGTGGGTGAAGACATCTG	AGAAATGTTGGCAAG	660
Qy	661	AAGTGGGATCT	CCAGATGATC	AGAAACATTTG	ATAGACAAATTTTATAGAA	720
Db	661	AAGTGGGATCT	CCAGATGATC	AGAAACATTTG	ATAGACAAATTTTATAGAA	720
Qy	721	TGGTCAAAAA	GAACAAAGACAC	CAATTACATAC	TAGTAGTTGGTGGCTACAGAAAAGCT	780
Db	721	TGGTCAAAAA	GAACAAAGACAC	CAATTACATAC	TAGTAGTTGGTGGCTACAGAAAAGCT	780
Qy	781	GACCTTGAAT	CAAGATATAAT	TGCACTCTTTTGT	CTGGCGAAATGGAGAAAAGAGA	840
Db	781	GACCTTGAAT	CAAGATATAAT	TGCACTCTTTTGT	CTGGCGAAATGGAGAAAAGAGA	840
Qy	841	TGGACCTGG	CGCTACTCGGAAATTTTGT	CTCGCGTTTGGTGTGTTAT	TGATCTCGAGAAA	900
Db	841	TGGACCTGG	CGCTACTCGGAAATTTTGT	CTCGCGTTTGGTGTGTTAT	TGATCTCGAGAAA	900
Qy	901	ACCTCGAAT	TAAGATCTTTTGA	ACCTTTAATAG	SAGAGATACCATTTGAATA	954
Db	901	ACCTCGAAT	TAAGATCTTTTGA	ACCTTTAATAG	SAGAGATACCATTTGAATA	954

RESULT 2

```

1 Sequence 1, Application US/08518615A
2 Patent No. 5962258
3
4 GENERAL INFORMATION:
5 APPLICANT: Mathur, E., et al.
6 TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
7 NUMBER OF SEQUENCES: 4
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
10 ADDRESSEE: CECCHI, STEWART & OLSTEIN
11 STREET: 6 BECKER FARM ROAD
12 CITY: ROSELAND
13 STATE: NEW JERSEY
14 COUNTRY: USA
15
16 ZIP: 07068
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: 3.5 INCH DISKETTE
20 COMPUTER: IBM PS/2
21 OPERATING SYSTEM: MS-DOS
22 SOFTWARE: WORD PERFECT 5.1
23
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/08/518,615A
26 FILING DATE: August 23, 1995
27
28 CLASSIFICATION: 435
29
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER:
32 FILING DATE:
33
34 ATTORNEY/AGENT INFORMATION:
35 NAME: FERRARO, GREGORY D.
36 REGISTRATION NUMBER: 36,134
37 REFERENCE/DOCKET NUMBER: 331400-20
38 TELECOMMUNICATION INFORMATION:
39 TELEPHONE: 201-994-1700
40 TELEFAX: 201-994-1744
41
42 INFORMATION FOR SEQ ID NO: 1:
43
44 SEQUENCE CHARACTERISTICS:
45 LENGTH: 954 BASE PAIRS
46 TYPE: NUCLEIC ACID
47 STRANDEDNESS: SINGLE
48 TOPOLOGY: LINEAR
49
50 MOLECULE TYPE: CDNA
51
52 US-08-518-615A-1

```

Query Match	100.0%;	Score 954;	DB 2;	Length 954;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 954;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0.0%;

Qy	1	ATGSGTGTGATCCCTTTTGGAAAGAACCAAAATTTGGGAAAGGCACTTAATATAGAAAT	60
Db	1	ATGGGTGTGTATCCTTTTGAAGGACAAAATATTGGAGAGGCACTTAATATAGAAAT	60
Qy	61	GCGCTTGAGACCAAAATGAGGGAGACTGGGGAGTGGTGTAAAGATGATGTTCCGAC	120
Db	61	GCGCTTGAGACCAAAATGAGGGAGACTGGGGAGTGGTGTAAAGATGATGTTCCGAC	120
Qy	121	ATTATAAAAGACCGGTTTCTCATGTTCCGAATTCCAATTAAGATGAGGTACGACGCT	180
Db	121	ATTATAAAAGACCGGTTTCTCATGTTCCGAATTCCAATTAAGATGAGGTACGACGCT	180
Qy	181	TACGGTTCCTCCTTATAAATCATGGATGGCTTCCTCAAAAGAGTGAAGTACGATA	240
Db	181	TACGGTTCCTCCTTATAAATCATGGATGGCTTCCTCAAAAGAGTGAAGTACGATA	240
Qy	241	AACGAGCCCTGMAAAGAGAGTGGCTGTGTATAAATATTCATCATAAGAGAGTTA	300
Db	241	AACGAGCCCTGMAAAGAGAGTGGCTGTGTATAAATATTCATCATAAGAGAGTTA	300
Qy	301	ATGAATGATCCAGAGAACACACAGAGAAAGATTCTGCTCTTTGGAAACAAATTGTGTAT	360
Db	301	ATGATATGATCCAGAGAACACACAGAGAAAGATTCTGCTCTTTGGAAACAAATTGTGTAT	360
Qy	361	GCTTATAAAGCTATCCCGGAACCTATTTTTTCAAAATTCGATGAACCTCCGAGAAAT	420
Db	361	GCTTATAAAGCTATCCCGGAACCTATTTTTTCAAAATTCGATGAACCTCCGAGAAAT	420
Qy	421	CTTACTCCGGAANAATGGAATGAACCTGCTGTAGGAGACCTCTAANAAGTTATAGATCAATT	480
Db	421	CTTACTCCGGAANAATGGAATGAACCTGCTGTAGGAGACCTCTAANAAGTTATAGATCAATT	480
Qy	481	GACAAAAGACACACTATATATATATAGCACACTGAATGGGGGTATATCTGCGCTTGAA	540
Db	481	GACAAAAGACACACTATATATATATAGCGCACACTGAATGGGGGTATATCTGCGCTTGAA	540
Qy	541	AAACGTGTCGTCCCAAAATGGGAAAAAAATTCCTATGTAACAATTCCTACTACAATCCY	600
Db	541	AAACGTGTCGTCCCAAAATGGGAAAAAAATTCCTATGTAACAATTCCTACTACAATCCY	600
Qy	601	TTTCGAATTTACCATCAAGAGAGCTGAGTGGTGAAGGATCTGGAANAATGTTGGGAGAGA	660
Db	601	TTTCGAATTTACCATCAAGAGAGCTGAGTGGTGAAGGATCTGGAANAATGTTGGGAGAGA	660
Qy	661	AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGAAAGAA	720
Db	661	AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGAAAGAA	720
Qy	721	TGCTCAAAAACAAACAAAGACCAATTTACATAGGTGAGTTTGGTGCCTACAGAAAAGCT	780
Db	721	TGCTCAAAAACAAACAAAGACCAATTTACATAGGTGAGTTTGGTGCCTACAGAAAAGCT	780
Qy	781	GACCTTGAATCAGAGATAAATAGSACCTCTTGTCTGCTCGCAAAATGAGAAAAAGAGA	840
Db	781	GACCTTGAATCAGAGATAAATAGSACCTCTTGTCTGCTCGCAAAATGAGAAAAAGAGA	840
Qy	841	TGAGCTGGGCATCTAGGGAATTTTGTTCGCGTTTGGTATTATAGACTCTGAGAAAA	900
Db	841	TGAGCTGGGCATCTAGGGAATTTTGTTCGCGTTTGGTATTATAGACTCTGAGAAAA	900
Qy	901	ACCTGGAATTAAGATCTTTTACAGCTTTAATAGGAGAGATAGACATTTGATTA	954
Db	901	ACCTGGAATTAAGATCTTTTACAGCTTTAATAGGAGAGATAGACATTTGATTA	954

RESULT 3

```

US-08-951-883-1
; Sequence 1, Application US/08951889
; Patent No. 6008032
; GENERAL INFORMATION:
; APPLICANT: Malthur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from

```

```

? TITLE OF INVENTION: Thermotoga Maritima
? NUMBER OF SEQUENCES: 4
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: CABELLA, BYRNE, BAIN, GILFILLAN,
? ADDRESSEE: CECCHI, STEWART & OLSTEIN
? STREET: 6 BECKER FARM ROAD
? CITY: ROSELAND
? STATE: NEW JERSEY
? COUNTRY: USA
? ZIP: 07068
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5 INCH DISKETTE
? COMPUTER: IBM PS/2
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: WORD PERFECT 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/951,889
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/518,615
? FILING DATE: August 23, 1995
? ATTORNEY/AGENT INFORMATION:
? NAME: FERRARO, GREGORY D.
? REGISTRATION NUMBER: 36,134
? REFERENCE/DOCKET NUMBER: 331400-20
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 201-994-1700
? TELEFAX: 201-994-1744
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 954 BASE PAIRS
? TYPE: NUCLEIC ACID
? STRANDEDNESS: SINGLE
? TOPOLOGY: LINEAR
? MOLECULE TYPE: CDNA
? US-08-951-889-1

Query Match 100.0%; Score 954; DB 3; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCTTGATCCTTTGGAAGAACAAATATGGAAGAGCATTAATATAGGAAT 60
DB 1 ATGGGCTTGATCCTTTGGAAGAACAAATATGGAAGAGCATTAATATAGGAAT 60
QY 61 GCGCTTGAGGACCAATGAGGAGACTGGGAGTGTATTAAGATGATCTTCGAC 120
DB 61 GCGCTTGAGGACCAATGAGGAGACTGGGAGTGTATTAAGATGATCTTCGAC 120
QY 121 ATATATAAAGAGCCGGTTCTCATGTTCGAATTCATAAGATGATGACGCGCT 180
DB 121 ATATATAAAGAGCCGGTTCTCATGTTCGAATTCATAAGATGATGACGCGCT 180
QY 181 TACGCTTTCCTCTATATAATCATGATCGCTTCTCAAAAGAGTGGATGATGATA 240
DB 181 TACGCTTTCCTCTATATAATCATGATCGCTTCTCAAAAGAGTGGATGATGATA 240
QY 241 AAGGAGCCCTGAAAAGAGAGCTGGCTGCTATTAATTCATACATGACGAGCTTA 300
DB 241 AAGGAGCCCTGAAAAGAGAGCTGGCTGCTATTAATTCATACATGACGAGCTTA 300
QY 301 ATGAATGATCCAGAAAGACACAGAAAGATTTCTGCTTTGGAACAAATTCGTGAT 360
DB 301 ATGAATGATCCAGAAAGACACAGAAAGATTTCTGCTTTGGAACAAATTCGTGAT 360
QY 361 CGTTATAAAGACTATCCGAAACTGATTTTGAATTCGATGATGACGCGAAAT 420
DB 361 CGTTATAAAGACTATCCGAAACTGATTTTGAATTCGATGATGACGCGAAAT 420
QY 421 CTTACTCGGAAAAATGAAATGAACTGTTGAGGAGCTCTAAAGTTTAAATGATCAAT 480
DB 421 CTTACTCGGAAAAATGAAATGAACTGTTGAGGAGCTCTAAAGTTTAAATGATCAAT 480
```

```

QY 481 GACAAAAGCACATATATATATAGGCACAGCTGAATGGGGGGTATATCTGCCCTTGA 540
DB 481 GACAAAAGCACATATATATATAGGCACAGCTGAATGGGGGGTATATCTGCCCTTGA 540
QY 541 AAACGTCTGTCCCAAAATGGGAAAAAATTCATAGTTACAAATTCACATCAATCC 600
DB 541 AAACGTCTGTCCCAAAATGGGAAAAAATTCATAGTTACAAATTCACATCAATCC 600
QY 601 TTGGAATTTACCATCAGAGAGCTGGCTGGAGAGATCTAGAAAATGGTTGGAGAG 660
DB 601 TTGGAATTTACCATCAGAGAGCTGGCTGGAGAGATCTAGAAAATGGTTGGAGAG 660
QY 661 AAGTGGGATCTCCAGATGATCAGAAAACATTTGATGAGAATTCATTTATAGAAGA 720
DB 661 AAGTGGGATCTCCAGATGATCAGAAAACATTTGATGAGAATTCATTTATAGAAGA 720
QY 721 TGGTCAAAAAAGAACAAAGACCAATTTACATAGTGAGTTGGTGCCTACAGAAAGCT 780
DB 721 TGGTCAAAAAAGAACAAAGACCAATTTACATAGTGAGTTGGTGCCTACAGAAAGCT 780
QY 781 GACCTGAATCAGAAATPAAATGAGACCTCCTTGTGTCGGAATGGAAGAGAGAG 840
DB 781 GACCTGAATCAGAAATPAAATGAGACCTCCTTGTGTCGGAATGGAAGAGAGAG 840
QY 841 TGGAGCTGGGCATCTGGAATTTTTCGGTTTGGTGTATGATCTGAGAGAGAG 900
DB 841 TGGAGCTGGGCATCTGGAATTTTTCGGTTTGGTGTATGATCTGAGAGAGAG 900
QY 901 ACCTGGAATTAAGATCTTTTGAAGCTTTAATAGGAGAGATGACATGTAATA 954
DB 901 ACCTGGAATTAAGATCTTTTGAAGCTTTAATAGGAGAGATGACATGTAATA 954
```

```

RESULT 4
US-09-472-857-1
? Sequence 1, Application US/09472857
? Patent No. 6245547
? GENERAL INFORMATION:
? APPLICANT: Mathur, E., et al.
? TITLE OF INVENTION: Carboxymethyl Cellulase from
? NUMBER OF SEQUENCES: 4
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: CABELLA, BYRNE, BAIN, GILFILLAN,
? ADDRESSEE: CECCHI, STEWART & OLSTEIN
? STREET: 6 BECKER FARM ROAD
? CITY: ROSELAND
? STATE: NEW JERSEY
? COUNTRY: USA
? ZIP: 07068
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5 INCH DISKETTE
? COMPUTER: IBM PS/2
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: WORD PERFECT 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/472,857
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/951,889
? FILING DATE:
? APPLICATION NUMBER: 08/518,615
? FILING DATE: August 23, 1995
? ATTORNEY/AGENT INFORMATION:
? NAME: FERRARO, GREGORY D.
? REGISTRATION NUMBER: 36,134
? REFERENCE/DOCKET NUMBER: 331400-20
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 201-994-1700
? TELEFAX: 201-994-1744
? INFORMATION FOR SEQ ID NO: 1:
```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 954 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
; US-09-472-857-1

```

```

Query Match          100.0%; Score 954; DB 4; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ATGGGTGTTGATCCTTTTGAAGAACAATAATTTGGGAGAGCGATTATATAGGAAT 60
DB 1 ATGGGTGTTGATCCTTTTGAAGAACAATAATTTGGGAGAGCGATTATATAGGAAT 60
QY 61 GCGGTTGAAGCACCATAATGAGGAGAGCTGGGTATGAAGATGATCTTTCGAC 120
DB 61 GCGGTTGAAGCACCATAATGAGGAGAGCTGGGTATGAAGATGATCTTTCGAC 120
QY 121 ATTATGAAGAAGCGCGTTTCTCATGTTTCAATTAAGATGAGTACGACGCT 180
DB 121 ATTATGAAGAAGCGCGTTTCTCATGTTTCAATTAAGATGAGTACGACGCT 180
QY 181 TACGGGTTTCCCTTATTAATTCATGATCGCTTCTCAAAAGAGTGAAGTGA 240
DB 181 TACGGGTTTCCCTTATTAATTCATGATCGCTTCTCAAAAGAGTGAAGTGA 240
QY 241 AACGGAGCCCTGAAGAAGAGAGTGGCTGTTGCTATTAATTCATCAGAGAGTTA 300
DB 241 AACGGAGCCCTGAAGAAGAGAGTGGCTGTTGCTATTAATTCATCAGAGAGTTA 300
QY 301 ATGAATGATCCAGAAAGACACAAGAAAGATTTCCTGCTTTTGGAAACAAATTCGTGAT 360
DB 301 ATGAATGATCCAGAAAGACACAAGAAAGATTTCCTGCTTTTGGAAACAAATTCGTGAT 360
QY 361 CGTTTAAAGATATCCCGCAAACTCATTTTGAATTCGATGAACCTCAGCAAAAT 420
DB 361 CGTTTAAAGATATCCCGCAAACTCATTTTGAATTCGATGAACCTCAGCAAAAT 420
QY 421 CTTACTCCGGAATAATGAGATGACTGCTTGAAGAGCTCTTAAGATCAAT 480
DB 421 CTTACTCCGGAATAATGAGATGACTGCTTGAAGAGCTCTTAAGATCAAT 480
QY 481 GCAAAAACACACATATATTAATGAGCAGAGTGAATGGGGGGTATATCTGCCCTTGA 540
DB 481 GCAAAAACACACATATATTAATGAGCAGAGTGAATGGGGGGTATATCTGCCCTTGA 540
QY 541 AAACGTCTGTCCCAAAATGGGAAAAAATTTCTATAGTTACAATTCACACTACATCTCT 600
DB 541 AAACGTCTGTCCCAAAATGGGAAAAAATTTCTATAGTTACAATTCACACTACATCTCT 600
QY 601 TTGCAATTTACCATATCAAGAGCTGAGTGGTGAAGAGATCAGAAATGGTGGGAGA 660
DB 601 TTGCAATTTACCATATCAAGAGCTGAGTGGTGAAGAGATCAGAAATGGTGGGAGA 660
QY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATGAAGAAATTCATTTTATAGAA 720
DB 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATGAAGAAATTCATTTTATAGAA 720
QY 721 TGGTCAAAAAGAAAGAACCAATTTACATAGTGTGTTGGTGCCTACAGAAAGCT 780
DB 721 TGGTCAAAAAGAAAGAACCAATTTACATAGTGTGTTGGTGCCTACAGAAAGCT 780
QY 781 GACCTTGAATCAAGATTAATATGAGCTCTTGTGCTTGGCAAAATGGGAAAGGAGA 840
DB 781 GACCTTGAATCAAGATTAATATGAGCTCTTGTGCTTGGCAAAATGGGAAAGGAGA 840
QY 841 TGGAGCTGGGATATGGAATTTTGTTCGGTTTGGTGTATGATCTGAGAAAA 900
DB 841 TGGAGCTGGGATATGGAATTTTGTTCGGTTTGGTGTATGATCTGAGAAAA 900
QY 901 ACCTGAATTAAGATCTTTTAGAAGCTTTAATAGGAGAGATGACATTGATAA 954

```

```

DB 901 ACCTGAATTAAGATCTTTTAGAAGCTTTAATAGGAGAGATGACATTGATAA 954

```

```

RESULT 5
US-09-066-075-3
; Sequence 3, Application US/09066075
; Patent No. 5925749
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,075
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/518,615
; FILING DATE: August 23, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: Oligonucleotide
; US-09-066-075-3
;
Query Match          2.2%; Score 21; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ATGGGTGTTGATCCTTTTGA 21
DB 31 ATGGGTGTTGATCCTTTTGA 51

```

```

RESULT 6
US-08-518-615A-3
; Sequence 3, Application US/08518615A
; Patent No. 5962258
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068

```

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/518,615A
FILING DATE: August 23, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: Oligonucleotide
US-08-518-615a-3

Query Match 2.2%; Score 21; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGGTGTGATCCTTTGAA 21
|||||
DB 31 ATGGGTGTGATCCTTTGAA 51

RESULT 7
US-08-951-889-3
Sequence 3, Application US/08951889
Patent No. 6008032
GENERAL INFORMATION:
APPLICANT: Mathur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from
TITLE OF INVENTION: Thermotoga Maritima
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951,889
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: Oligonucleotide
US-08-951-889-3

Query Match 2.2%; Score 21; DB 3; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGGTGTGATCCTTTGAA 21
|||||
DB 31 ATGGGTGTGATCCTTTGAA 51

RESULT 8
US-09-472-857-3
Sequence 3, Application US/09472857
Patent No. 6245547
GENERAL INFORMATION:
APPLICANT: Mathur, E., et al.
TITLE OF INVENTION: Carboxymethyl Cellulase from
TITLE OF INVENTION: Thermotoga Maritima
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/472,857
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/951,889
FILING DATE:
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: Oligonucleotide
US-09-472-857-3

Query Match 2.2%; Score 21; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGGGTGTGATCCTTTGAA 21
|||||
DB 31 ATGGGTGTGATCCTTTGAA 51

Mon Apr 28 09:31:58 2003

us-09-880-729a-1.oli20.rni

Page 6

Search completed: April 26, 2003, 15:27:26
job time : 47 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 11:55:53 ; Search time 98 Seconds
(without alignments)
10592.600 Million cell updates/sec

Title: US-09-880-729A-1
Perfect score: 954
Sequence: 1 atgggtgttgatccttga.....gagagatagcatgataa 954

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 709820 seqs, 544064369 residues

Word size : 20

Total number of hits satisfying chosen parameters: 6

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

Published Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	954	100.0	954	US-09-880-729-1	Sequence 1, Appl1
2	21	2.2	51	US-09-880-729-3	Sequence 3, Appl1
3	20	2.1	530	US-09-764-877-824	Sequence 824, Appl1
4	20	2.1	1350	US-09-764-877-3648	Sequence 3648, Appl1
5	20	2.1	1350	US-09-764-877-3649	Sequence 3649, Appl1
6	20	2.1	1350	US-09-764-877-3650	Sequence 3650, Appl1

ALIGNMENTS

RESULT 1
US-09-880-729-1
; Sequence 1, Application US/09880729
; Publication No. US20030044956A1
; GENERAL INFORMATION:
; APPLICANT: Mathur, E., et al.
; TITLE OF INVENTION: Carboxymethyl Cellulase from
; Thermotoga Maritima

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/880,729
FILING DATE: 12-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/472,857
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/951,889
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/518,615
FILING DATE: August 23, 1995
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 331400-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-880-729-1
Query Match 100.0%; Score 954; DB 9; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGGGTGTGATCCTTTGGAAGAACAAATATTTGGGAAGGACATTAATATAGCAAT 60
1 ATGGGTGTGATCCTTTGGAAGAACAAATATTTGGGAAGGACATTAATATAGCAAT 60
61 GCGCTTGAAGCACCACCAATGAGGAGACTGGGAGTGTGTAAGAGTGAAGTCTTGAC 120
61 GCGCTTGAAGCACCACCAATGAGGAGACTGGGAGTGTGTAAGAGTGAAGTCTTGAC 120
121 ATATATAAAGAACCCGTTCTCTCATGTTGGAATTCACATTAAGATGAGTACGACGT 180
121 ATATATAAAGAACCCGTTCTCTCATGTTGGAATTCACATTAAGATGAGTACGACGT 180
181 TACGCGTTCTCTCTATATAAATCATGATGCTTCTTCAAAAGAGTGAAGTGATA 240
181 TACGCGTTCTCTCTATATAAATCATGATGCTTCTTCAAAAGAGTGAAGTGATA 240
241 AACGAGCCCTGAAAGAGAGCTGCTGTATATAATATGATCATGCTGAGAGATTA 300
241 AACGAGCCCTGAAAGAGAGCTGCTGTATATAATATGATCATGCTGAGAGATTA 300
301 ATGAATGATCCAGAGAACCAAGGAAGATTTCTGCTTTGGAAGCAATTTGCTGAT 360
301 ATGAATGATCCAGAGAACCAAGGAAGATTTCTGCTTTGGAAGCAATTTGCTGAT 360
361 CGTTATTAAGACTATCCCGAAGCTATTTTGAAGTTTGAATGATGACCTCAGCAAT 420
361 CGTTATTAAGACTATCCCGAAGCTATTTTGAAGTTTGAATGATGACCTCAGCAAT 420

```

421 CTCTACCGAGAAATGSAATGAACCTGCTTGAAGAGCTTAAGATATAGATCAAT 480
421 CTTACTCGGAAATGAATGAATGCTTGTGAGAGCTCTAAAGTTATAGATCAAT 480
481 GACAAAAGCAGCTATATATAGGAGCAGCTGAATGGGGGGCTATATCCCTTGA 540
481 GACAAAAGCAGCTATATATAGGAGCAGCTGAATGGGGGGCTATATCCCTTGA 540
541 AAAGCTGTGCTCCAAATGGGAAAAATCTATAGTTACCAATCAGTACTACATCT 600
541 AAAGCTGTGCTCCAAATGGGAAAAATCTATAGTTACCAATCAGTACTACATCT 600
601 TTGCAATTTACCATCAAGAGCTGAGTGGTGAAGAGCTGAGCAATGCTGGGAGA 660
601 TTGCAATTTACCATCAAGAGCTGAGTGGTGAAGAGCTGAGCAATGCTGGGAGA 660
661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGAGA 720
661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGAGA 720
721 TGGTCAAAAAGAACAAAGACCAATTTACATAGGTGAGTTGCTTACAGAAAAGCT 780
721 TGGTCAAAAAGAACAAAGACCAATTTACATAGGTGAGTTGCTTACAGAAAAGCT 780
781 GACCTGAATCAAGATTAATAGACCTCTTCTGCTTCCGGAATGAGAGAAAAGAGA 840
781 GACCTGAATCAAGATTAATAGACCTCTTCTGCTTCCGGAATGAGAGAAAAGAGA 840
841 TGGACCTGGGCGCTACTGGGAATTTTGTCCGTTTGTGTTATGATCTCTGAGAAAA 900
841 TGGACCTGGGCGCTACTGGGAATTTTGTCCGTTTGTGTTATGATCTCTGAGAAAA 900
901 ACCTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGGAGATAGCATTAATA 954
901 ACCTGGAATTAAGATCTTTAGAGCTTTAATAGAGAGGAGATAGCATTAATA 954

```

RESULT 2

```

US-09-880-729-3
; Sequence 3, Application US/09880729
; Publication No. US20030044956A1

```

GENERAL INFORMATION:

```

; APPLICANT: Mathur, E., et al.

```

```

; TITLE OF INVENTION: Carboxymethyl Cellulase from
; Thermotoga Maritima

```

```

; NUMBER OF SEQUENCES: 4

```

```

; CORRESPONDENCE ADDRESS:
; CARELLA, BYRNE, BAIN, GILFILLAN,
; CECCHI, STEWART & OLSTEIN

```

```

; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND

```

```

; STATE: NEW JERSEY
; COUNTRY: USA

```

```

; ZIP: 07068

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE

```

```

; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS

```

```

; SOFTWARE: WORD PERFECT 5.1

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/880,729

```

```

; FILING DATE: 12-Jun-2001

```

```

; CLASSIFICATION: <Unknown>

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/472,857

```

```

; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/951,889

```

```

; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/518,615

```

```

; FILING DATE: August 23, 1995

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.

```

```

; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 331400-20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: Oligonucleotide
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-880-729-3

```

```

Query Match 2.2%; Score 21; DB 9; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 ATGGGTGTGATCCCTTTGAA 21
Db 31 ATGGGTGTGATCCCTTTGAA 51

```

RESULT 3

```

US-09-764-877-824
; Sequence 824, Application US/09764877
; Patent No. US20020147140A1

```

GENERAL INFORMATION:

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

```

```

; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877

```

```

; PRIOR APPLICATION DATA: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper

```

```

; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 824
; LENGTH: 530

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens

```

```

US-09-764-877-824

```

```

Query Match 2.1%; Score 20; DB 10; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 891 TCTGAGAAAACCTGGATA 910
Db 63 TCTGAGAAAACCTGGATA 82

```

RESULT 4

```

US-09-764-877-3648
; Sequence 3648, Application US/09764877
; Patent No. US20020147140A1

```

GENERAL INFORMATION:

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

```

```

; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877

```

```

; PRIOR APPLICATION DATA: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper

```

```

; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 3648
; LENGTH: 1350

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens

```

```

US-09-764-877-3648

```

```

Query Match 2.1%; Score 20; DB 10; Length 1350;
Best Local Similarity 100.0%; Pred. No. 2.8;

```

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 891 TCTGAGAAAAACCTGGAATA 910
 |||
 Db 866 TCTGAGAAAAACCTGGAATA 885

RESULT 5

US-09-764-877-3649
 ; Sequence 3649, Application US/09764877
 ; Patent No. US20020147140A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC005
 ; CURRENT APPLICATION NUMBER: US/09/764,877
 ; Prior application data removed - refer to PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 4031
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO 3649
 ; LENGTH: 1350
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-877-3649

Query Match

2.1%; Score 20; DB 10; Length 1350;

Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 891 TCTGAGAAAAACCTGGAATA 910
 |||
 Db 866 TCTGAGAAAAACCTGGAATA 885

RESULT 6

US-09-764-877-3650
 ; Sequence 3650, Application US/09764877
 ; Patent No. US20020147140A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC005
 ; CURRENT APPLICATION NUMBER: US/09/764,877
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - refer to PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 4031
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO 3650
 ; LENGTH: 1350
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-877-3650

Query Match

2.1%; Score 20; DB 10; Length 1350;

Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 891 TCTGAGAAAAACCTGGAATA 910
 |||
 Db 866 TCTGAGAAAAACCTGGAATA 885

Search completed: April 26, 2003, 16:05:07
 Job time : 100 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 10:49:02 ; Search time 1910 Seconds
(without alignments)
14536.166 Million cell updates/sec

Title: US-09-880-729a-1

Perfect score: 954
Sequence: 1 atgggtgtgtatcccttttga.....gagagatagatcattgataa 954

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 20

Total number of hits satisfying chosen parameters: 87

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

Genembi:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_pro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pin:*
- 35: em_htg_rtd:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	954	100.0	954	6	AR077712	AR077712 Sequence
2	954	100.0	954	6	AR096552	AR096552 Sequence
3	954	100.0	954	6	AR157661	AR157661 Sequence
4	903	94.7	10244	1	AE001813	AE001813 Thermotog
5	23	2.4	150882	2	AC079067	AC079067 Homo sapi
6	23	2.4	160529	9	AC092659	AC092659 Homo sapi
7	23	2.3	178997	2	AC126611	AC126611 Mus muscu
8	22	2.3	194378	2	AC108661	AC108661 Rattus no
9	21	2.2	51	6	AR077713	AR077713 Sequence
10	21	2.2	51	6	AR096553	AR096553 Sequence
11	21	2.2	51	6	AR157662	AR157662 Sequence
12	21	2.2	106571	9	HS86E14	299572 Human DNA s
13	21	2.2	134757	9	AC007870	AC007870 Genomtic s
14	21	2.2	134781	10	AL355177	AL355177 Mouse DNA
15	21	2.2	137154	9	AL513082	AL513082 Human DNA
16	21	2.2	140795	2	AC079562	AC079562 Mus muscu
17	21	2.2	146385	9	AC010534	AC010534 Homo sapi
18	21	2.2	166496	9	AC124046	AC124046 Homo sapi
19	21	2.2	175518	9	AC018846	AC018846 Homo sapi
20	21	2.2	187038	2	AC021179	AC021179 Homo sapi
21	21	2.2	196216	2	AC019228	AC019228 Homo sapi
22	21	2.2	216988	10	AC084217	AC084217 Mus Muscu
23	21	2.2	243381	2	AC125205	AC125205 Mus Muscu
24	20	2.1	849	6	E06193	E06193 CDNA encodl
25	20	2.1	849	6	E06204	E06204 CDNA encodl
26	20	2.1	849	6	E06205	E06205 CDNA encodl
27	20	2.1	849	6	E06206	E06206 CDNA encodl
28	20	2.1	849	6	E06389	E06389 CDNA encodl
29	20	2.1	849	6	E06400	E06400 CDNA encodl
30	20	2.1	849	6	E06401	E06401 CDNA encodl
31	20	2.1	849	6	E06402	E06402 CDNA encodl
32	20	2.1	1280	6	E06200	E06200 CDNA encodl
33	20	2.1	1280	6	E06216	E06216 CDNA encodl
34	20	2.1	1280	6	E06217	E06217 CDNA encodl
35	20	2.1	1280	6	E06396	E06396 CDNA encodl
36	20	2.1	1280	6	E06412	E06412 CDNA encodl
37	20	2.1	1280	6	E06413	E06413 CDNA encodl
38	20	2.1	3564	6	E06203	E06203 CDNA encodl
39	20	2.1	3564	6	E06223	E06223 CDNA encodl
40	20	2.1	3564	6	E06399	E06399 CDNA encodl
41	20	2.1	3564	6	E06419	E06419 CDNA encodl
42	20	2.1	9359	14	AF313916	AF313916 Hepatitis C
43	20	2.1	9431	14	HPCHCPO	D45172 Hepatitis C
44	20	2.1	9431	14	HPCHCPO	D30613 Hepatitis C
45	20	2.1	9471	6	E06261	E06261 CDNA encodl
46	20	2.1	9471	6	E06457	E06457 CDNA encodl
47	20	2.1	9504	14	AB049099	AB049099 Hepatitis
48	20	2.1	10435	1	AE008634	AE008634 Ricketts
49	20	2.1	13610	1	AE013181	AE013181 Thermoana
50	20	2.1	39726	3	CEB0413	Z92824 Caenorhabdi
51	20	2.1	46206	9	AL591062	AL591062 Human DNA
52	20	2.1	55577	2	AC014109	AC014109 Drosophila
53	20	2.1	80025	9	AL391355	AL391355 Human DNA
54	20	2.1	84281	2	AC097840	AC097840 Rattus no
55	20	2.1	104753	2	AC114062	AC114062 Rattus no
56	20	2.1	109922	2	AC011671	AC011671 Homo sapi
57	20	2.1	110000	2	CEY3986_2	Continuation (3 of
58	20	2.1	117000	9	AC069222	AC069222 Homo sapi
59	20	2.1	117728	8	AC006586	AC006586 Arabidops
60	20	2.1	126547	2	AC103198	AC103198 Rattus no
61	20	2.1	139571	2	AC095268	AC095268 Rattus no
62	20	2.1	144544	8	AC119071	AC119071 Oryza sat
63	20	2.1	151281	9	AC103660	AC103660 Homo sapi
64	20	2.1	152233	2	AF268619	AF268619 Homo sapi
65	20	2.1	155204	2	AC007926	AC007926 Trypanoso

Pred. No. is the number of results predicted by chance to have a

```

c 66 20 2.1 155761 2 AP005578
c 67 20 2.1 158774 2 AP005587
c 68 20 2.1 159020 2 AC092014
c 69 20 2.1 159356 2 AL845427
c 70 20 2.1 162458 2 AC128759
c 71 20 2.1 166991 2 AC015952
c 72 20 2.1 166992 2 AC022883
c 73 20 2.1 168366 9 AC010002
c 74 20 2.1 169921 9 AC093497
c 75 20 2.1 170071 9 AC092022
c 76 20 2.1 170141 9 AC091559
c 77 20 2.1 170423 2 AC092787
c 78 20 2.1 171044 2 AC095222
c 79 20 2.1 174974 2 AC119585
c 80 20 2.1 179678 2 AC123247
c 81 20 2.1 198043 2 AC123247
c 82 20 2.1 217766 10 AC087558
c 83 20 2.1 222368 2 AC103652
c 84 20 2.1 246237 3 CE93986A
c 85 20 2.1 264211 2 AC123867
c 86 20 2.1 274349 2 AC093623
c 87 20 2.1 295377 3 AE003532
```

ALIGNMENTS

```

RESULT 1
LOCUS AR077712 954 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 1 from patent US 5962258.
ACCESSION AR077712.1 GI:10004458
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 954)
AUTHORS Mathur,E.J. and Lam,D.E.
TITLE Carboxymethyl cellulase from thermotoga maritima
JOURNAL Patent: US 5962258-A 1 05-OCT-1999;
FEATURES
location/Qualifiers
source
```

```

BASE COUNT 335 a 145 c 221 g 253 t
ORIGIN
```

```

Query Match 100.0%; Score 954; DB 6; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 1 ATGGGTTGATGCTTTTGAAGAACAAATTTGGAGAGGCAATTAATAGAAAT 60
DB 1 ATGGGTTGATGCTTTTGAAGAACAAATTTGGAGAGGCAATTAATAGAAAT 60
QY 61 GCGCTTGAGCACCAGCAATGAGGAGTGGAGTGGTGAATAAAGATGAGTCTTGAC 120
DB 61 GCGCTTGAGCACCAGCAATGAGGAGTGGAGTGGTGAATAAAGATGAGTCTTGAC 120
QY 121 ATTATTAAGAGCGGTTTCTCTCATGTTGCAATTCATTAAGATGAGTAGCAGCT 180
DB 121 ATTATTAAGAGCGGTTTCTCTCATGTTGCAATTCATTAAGATGAGTAGCAGCT 180
QY 181 TAGCGTTTCTCTCTTAATAATCATGATCGCTTCTTCAAAAGAGTGAAGTATA 240
DB 181 TAGCGTTTCTCTCTTAATAATCATGATCGCTTCTTCAAAAGAGTGAAGTATA 240
QY 241 AAGGAGCCCTGANAAGAGAGCTGCTGTGCTATAATATTCATCACTACAGAGATTA 300
DB 241 AAGGAGCCCTGANAAGAGAGCTGCTGTGCTATAATATTCATCACTACAGAGATTA 300
QY 301 ATGATGATGCAGAGAGACACAGAGAAAGATTTTGTGCTTTTGGAAACAATTCGAT 360
DB 301 ATGATGATGCAGAGAGACACAGAGAAAGATTTTGTGCTTTTGGAAACAATTCGAT 360
```

```

DB 301 ATGATGATGCAGAGAGACACAGAGAAAGATTTTGTGCTTTTGGAAACAATTCGAT 360
QY 361 CGTTTAAAGACATATCCCGAAGCTCTATTTTGAATTTGATGATGAACCTCAGGAAT 420
DB 361 CGTTTAAAGACATATCCCGAAGCTCTATTTTGAATTTGATGATGAACCTCAGGAAT 420
QY 421 CTCTACCGGAAAATGGAATGAATGCTGTGAGAGAGCTTAAAGTTATAGATCAAT 480
DB 421 CTCTACCGGAAAATGGAATGAATGCTGTGAGAGAGCTTAAAGTTATAGATCAAT 480
QY 481 GACAAAAGACACATATATATATAGGACACAGTGAATGAGGGGCTATATCTGCCCTTGA 540
DB 481 GACAAAAGACACATATATATATAGGACACAGTGAATGAGGGGCTATATCTGCCCTTGA 540
QY 541 AAGCTGCTGCTCCCAATATGGAATAAATCTATAGTTACAAATTCACATCAATCCT 600
DB 541 AAGCTGCTGCTCCCAATATGGAATAAATCTATAGTTACAAATTCACATCAATCCT 600
QY 601 TTGCAATTTACCATCAAGAGAGCTGAGTGGTGAAGGATCGAANAATGGTTGGAGAA 660
DB 601 TTGCAATTTACCATCAAGAGAGCTGAGTGGTGAAGGATCGAANAATGGTTGGAGAA 660
QY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATGAGAAGATTCATTTATGAGAA 720
DB 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATGAGAAGATTCATTTATGAGAA 720
QY 721 TGGTCAAAAAGACAAAAGACCAATTTACATAGCTGAGTTGCTCTACAGAAAGCT 780
DB 721 TGGTCAAAAAGACAAAAGACCAATTTACATAGCTGAGTTGCTCTACAGAAAGCT 780
QY 781 GACCTTGAATCAAGATTAATGACCTCTTGTGCTGCTTCCGGAATGCAAAAAGAGA 840
DB 781 GACCTTGAATCAAGATTAATGACCTCTTGTGCTGCTTCCGGAATGCAAAAAGAGA 840
QY 841 TGGAGCTGGGCACTACTGGGAATTTTGTCCGCTTTGCTTTATGATACCTGAGAAA 900
DB 841 TGGAGCTGGGCACTACTGGGAATTTTGTCCGCTTTGCTTTATGATACCTGAGAAA 900
QY 901 ACCTGGAATTAAGATCTTTTGAAGCAATTTAATAGAGAGATGATGATTAAT 954
DB 901 ACCTGGAATTAAGATCTTTTGAAGCAATTTAATAGAGAGATGATGATTAAT 954
```

```

RESULT 2
LOCUS AR096552 954 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 1 from patent US 6008032.
ACCESSION AR096552
VERSION AR096552.1 GI:10025457
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 954)
AUTHORS Mathur,E.J. and Lam,D.E.
TITLE Carboxymethyl cellulase from Thermotoga maritima
JOURNAL Patent: US 6008032-A 1 28-DEC-1999;
FEATURES
location/Qualifiers
source
```

```

BASE COUNT 335 a 145 c 221 g 253 t
ORIGIN
Query Match 100.0%; Score 954; DB 6; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 1 ATGGGTTGATGCTTTTGAAGAACAAATTTGGAGAGGCAATTAATAGGAAT 60
DB 1 ATGGGTTGATGCTTTTGAAGAACAAATTTGGAGAGGCAATTAATAGGAAT 60
QY 61 GCGCTTGAGCACCAGCAATGAGGAGTGGAGTGGTGAATAAAGATGAGTCTTGAC 120
DB 61 GCGCTTGAGCACCAGCAATGAGGAGTGGAGTGGTGAATAAAGATGAGTCTTGAC 120
```

Db 61 GCCCTTGAGACCAAAATGAGAGACTGGGAGGTGATATAAGATGACTTCTGCAC 120
QY 121 ATTATAAAGAACCCGGTTTCTCTCATGTTCGAATTCATAAGATGAGATAGCAGCGT 180
Db 121 ATTATAAAGAACCCGGTTTCTCTCATGTTCGAATTCATAAGATGAGATAGCAGCGT 180
QY 181 TAGCGCTTCTCTCTTATATAATCATGATGCTTCTTCAAAAGAGTGAAGTGAATA 240
Db 181 TAGCGCTTCTCTCTTATATAATCATGATGCTTCTTCAAAAGAGTGAAGTGAATA 240
QY 241 AAGGAGCCCTGAAAAAGAGAGCTGGTGTGCTATATAATTTCTACTCTACGAGAGTTA 300
Db 241 AAGGAGCCCTGAAAAAGAGAGCTGGTGTGCTATATAATTTCTACTCTACGAGAGTTA 300
QY 301 ATGAATGATCCAGAAAGACCAAGAAAGATTTCTGCTTTGGAACCAATTTGCGAT 360
Db 301 ATGAATGATCCAGAAAGACCAAGAAAGATTTCTGCTTTGGAACCAATTTGCGAT 360
QY 361 CGTTATTAAGACTATCCCGAACTATTTTGGAAATTCGATGAACCTCAGCGAAT 420
Db 361 CGTTATTAAGACTATCCCGAACTATTTTGGAAATTCGATGAACCTCAGCGAAT 420
QY 421 CTATCTCCGAAAAATGAGATGAGCTGTGAGAGAGCTTAAAGTATTAAGATCAATT 480
Db 421 CTATCTCCGAAAAATGAGATGAGCTGTGAGAGAGCTTAAAGTATTAAGATCAATT 480
QY 481 GACAAAAAGACACTATTAATTAAGACAGCTGAATGGGGGTATATCTGCCCTTGA 540
Db 481 GACAAAAAGACACTATTAATTAAGACAGCTGAATGGGGGTATATCTGCCCTTGA 540
QY 541 AAACGTCTGTCCCAAAATGGGAAAAATTCATAGTTTCAATTCACACTACATCCT 600
Db 541 AAACGTCTGTCCCAAAATGGGAAAAATTCATAGTTTCAATTCACACTACATCCT 600
QY 601 TTGGAATTTACCATCAGAGAGCTGAGTGGTGAAGAGATCTGAAGATGGTTGGAGAA 660
Db 601 TTGGAATTTACCATCAGAGAGCTGAGTGGTGAAGAGATCTGAAGATGGTTGGAGAA 660
QY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGAAGAA 720
Db 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGAAGAA 720
QY 721 TGGTCAAAAAAGACAAAGACCAATTTACATAGTGTGAGTTGGTGCCTACAGAAAA 780
Db 721 TGGTCAAAAAAGACAAAGACCAATTTACATAGTGTGAGTTGGTGCCTACAGAAAA 780
QY 781 GACCTGGAATCAAGATTAATATGAGTGTGAGTTGGTGCCTACAGAAAA 840
Db 781 GACCTGGAATCAAGATTAATATGAGTGTGAGTTGGTGCCTACAGAAAA 840
QY 841 TGGAGCTGGGCATCTGAGCAATTTGTTCCGGTTTGGTGTATGATCTGAGAAAA 900
Db 841 TGGAGCTGGGCATCTGAGCAATTTGTTCCGGTTTGGTGTATGATCTGAGAAAA 900
QY 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAAATAGAGAGATAGCATTTGAATA 954
Db 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAAATAGAGAGATAGCATTTGAATA 954
RESULT 3
AR157661
LOCUS AR157661 954 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 1 from patent US 6245547.
ACCESSION AR157661
VERSION AR157661.1 GI:16218624
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 954)
AUTHORS Mathur E.J. and Lam D.E.
TITLE Carboxymethyl cellulose from thermotoga maritima
JOURNAL Patent: US 6245547-A I 12-JUN-2001;

FEATURES
source Location/Qualifiers
BASE COUNT 335 a 145 c 221 g 253 t
ORIGIN
Query Match 100.0%; Score 954; DB 6; Length 954;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGGTGTATCTCTTTGAAGAGACAAATATGGGAAGAGAGCATTAATATAGCAAT 60
Db 1 ATGGGTGTATCTCTTTGAAGAGACAAATATGGGAAGAGAGCATTAATATAGCAAT 60
QY 61 GCGCTTACACCAAAATGAGGAGAGTGGGAGTGGTGAATAAGATGAGTCTTGCAC 120
Db 61 GCGCTTACACCAAAATGAGGAGAGTGGGAGTGGTGAATAAGATGAGTCTTGCAC 120
QY 121 ATTATAAAGAACCGGTTTCTCTCATGTTCGAATTCATAAGATGAGATAGCAGCGT 180
Db 121 ATTATAAAGAACCGGTTTCTCTCATGTTCGAATTCATAAGATGAGATAGCAGCGT 180
QY 181 TAGCGCTTCTCTCTTATATAATCATGATGCTTCTTCAAAAGAGTGAAGTGAATA 240
Db 181 TAGCGCTTCTCTCTTATATAATCATGATGCTTCTTCAAAAGAGTGAAGTGAATA 240
QY 241 AAGGAGCCCTGAAAAAGAGAGCTGGTGTGCTATATAATTTCTACTCTACGAGAGTTA 300
Db 241 AAGGAGCCCTGAAAAAGAGAGCTGGTGTGCTATATAATTTCTACTCTACGAGAGTTA 300
QY 301 ATGAATGATCCAGAAAGACCAAGAAAGATTTCTGCTTTGGAACCAATTTGCGAT 360
Db 301 ATGAATGATCCAGAAAGACCAAGAAAGATTTCTGCTTTGGAACCAATTTGCGAT 360
QY 361 CGTTATTAAGACTATCCCGAACTATTTTGGAAATTCGATGAACCTCAGCGAAT 420
Db 361 CGTTATTAAGACTATCCCGAACTATTTTGGAAATTCGATGAACCTCAGCGAAT 420
QY 421 CTATCTCCGAAAAATGAGATGAGCTGTGAGAGAGCTTAAAGTATTAAGATCAATT 480
Db 421 CTATCTCCGAAAAATGAGATGAGCTGTGAGAGAGCTTAAAGTATTAAGATCAATT 480
QY 481 GACAAAAAGACACTATTAATTAAGACAGCTGAATGGGGGTATATCTGCCCTTGA 540
Db 481 GACAAAAAGACACTATTAATTAAGACAGCTGAATGGGGGTATATCTGCCCTTGA 540
QY 541 AAACGTCTGTCCCAAAATGGGAAAAATTCATAGTTTCAATTCACACTACATCCT 600
Db 541 AAACGTCTGTCCCAAAATGGGAAAAATTCATAGTTTCAATTCACACTACATCCT 600
QY 601 TTGGAATTTACCATCAGAGAGCTGAGTGGTGAAGAGATCTGAAGATGGTTGGAGAA 660
Db 601 TTGGAATTTACCATCAGAGAGCTGAGTGGTGAAGAGATCTGAAGATGGTTGGAGAA 660
QY 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGAAGAA 720
Db 661 AAGTGGGATCTCCAGATGATCAGAAACATTTGATAGAAATTCATTTTATAGAAGAA 720
QY 721 TGGTCAAAAAAGACAAAGACCAATTTACATAGTGTGAGTTGGTGCCTACAGAAAA 780
Db 721 TGGTCAAAAAAGACAAAGACCAATTTACATAGTGTGAGTTGGTGCCTACAGAAAA 780
QY 781 GACCTGGAATCAAGATTAATATGAGTGTGAGTTGGTGCCTACAGAAAA 840
Db 781 GACCTGGAATCAAGATTAATATGAGTGTGAGTTGGTGCCTACAGAAAA 840
QY 841 TGGAGCTGGGCATCTGAGCAATTTGTTCCGGTTTGGTGTATGATCTGAGAAAA 900
Db 841 TGGAGCTGGGCATCTGAGCAATTTGTTCCGGTTTGGTGTATGATCTGAGAAAA 900
QY 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAAATAGAGAGATAGCATTTGAATA 954
Db 901 ACCTGGAATTAAGATCTTTTGAAGCTTTTAAATAGAGAGATAGCATTTGAATA 954

RESULT 4
AE001813 10244 bp DNA linear BCT 02-JUN-1999
LOCUS Thermotoga maritima section 125 of 136 of the complete genome.
DEFINITION AE001813 AE000512
ACCESSION AE001813.1 GI:4982321
VERSION
KEYWORDS
SOURCE Thermotoga maritima.
ORGANISM Bacteria; Thermotogales; Thermotogales; Thermotogaceae;
Thermotoga.
REFERENCE
1 (bases 1 to 10244)
Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J.,
Hart, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A.,
McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M.,
Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A.,
Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D.,
Eisen, J.A., Fraser, C.M. et al.
Evidence for lateral gene transfer between Archaea and bacteria
from genome sequence of Thermotoga maritima
Nature 399 (6734), 323-329 (1999)
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
2 (bases 1 to 10244)
Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J.,
Hart, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A.,
McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M.,
Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A.,
Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D.,
White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 10244
/organism="Thermotoga maritima"
/db_xref="taxon:2336"
73. 1104
/gene="TM1744"
73. 1104
/gene="TM1744"
/note="similar to GB:U00096 SP:P75823 PID:1787095
PID:2225892 percent identity: 67.58; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="L-allo-threonine aldolase"
/protein_id="AAD36809.1"
/db_xref="GI:4982322"
/translation="MIDLSQVYTRKTEEMRKMAQAEVGDVYSGDPTINELERLAA
EFFGKALFVSGTGNQVSIWAHQGEVLEADSHIFREYEVAMVLSGVMPHP
VGGKNGAMPDDVRKAIIRPNHIFPTSLIAENTNRSGRVPLENTKEITAKE
HGIVNIDGARIFNASIAGVYADSVFQLSGKCAPSVVAGDFTIER
ARKARKMLGGMGAGVLAAGIATLKMDRLEKEDHENARFALKIKELGVSVDNED
VTKNVIILRTDNKIKVNAHGFIETLRNSGVLANVSDTEIRLIVTHKDVSRNDIEALNI
FEKLFRRKFS"
1183. 1806
/gene="TM1745"
1183. 1806
/gene="TM1745"
/note="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36810.1"
/db_xref="GI:4982323"
/translation="MKRFVVMVLITVALGVFAAGYGAERPAFEROPMGVQAQFPFC
CGGGLGIGPNAPMYERQMAERGFGLQIODEOLAKTRIAELNNLNDLD
DQIKELYNVAKETEOULREKRVQIHEEYINALVYKRDTKTAEIDKLDJDLRDEVQ
DIYKDFADKVDIITVDYRKRFGDKRMFLVLTDEGFVLEENVOE"

gene
CDS
2074. 4002
/gene="TM1746"
2074. 4002
/gene="TM1746"
/note="similar to GB:AE000783 percent identity: 56.54;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, periplasmic
oligopeptide-binding protein"
/protein_id="AAD36811.1"
/db_xref="GI:4982324"
/translation="MKRFVGMVLVLAAFSEYEVEVOLPMFGVTNFDPPYVQAQHL
ACGTFEGEGVYDPPKSLGKVVPAVCESTVSEDTITFTFLRKDRMSNDPPT
ALPPEFMKRAASPEPTPIPMASPVQYIKYVDDCKSGAKPLDELGVALDYLQVY
LSRPMSFTIMLVLGAMPLEHRTVOEHEDMMKPEYVNGPVYIESFTPNYEIVLY
RNNYVVDPEFGNDRKLVLRGGIGLQGYTAGELDVAETPAVGDPYFALKNNQSEIH
EIEAVVGIPEDEQAKKLAEAGYPNGKGFPRVFTYTGASDPVAELVDQMKVVG
ITPEIENIESGOSTLQMSDPVPMAPGCTTGGPVNODSGSLQOTADHFWFYDYP
AEYRQKLEYELREKRESYLREGGKTEEMKEILINRAQELMEKQVNSDEDFLAKY
FLPAPBELPFIQINKIYEMKKATTDOKIOLMRLAMRALIYNOEYSIVETGNNPT
NREARKLMARILSPYDKAVEYAKKIQQLCVDLAYEVPLVQKVVYLLKPYLGIVPY
KFSWGPTEFLFKYLIVK"
3999. 5024
/gene="TM1747"
3999. 5024
/gene="TM1747"
/note="similar to GB:U42023 SP:P45054 PID:1006439
PID:1221243 PID:1205369 percent identity: 65.90;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36812.1"
/db_xref="GI:4982325"
/translation="MNSPPDFGLLNNNGVVDMPFLKRLKLTLTIAISWVVIY
ITVYVLMAPGNEFELQVRDALARVTRTPDDPAVQAKTKEFERYGLNNPIKQILMY
LKGAVVFKEGFSFSDPARNIEDLKEFTPTFLALSLIFALVGVPIGLIALAKKN
TWIDYATMTVSVIGVAIPSYVAVFLLISYILGWPISGWSGIFRKLPTALAG
PLASVARTRVSLDITLNDQFIRVAVYKGGDDKRVIRKIALRSPMPLVYVGPQAY
LMGVGVWVENIFRIPLGQLFANAARVDYPLVTSFTLALVTVMNNLVDVLAAIL
DPRIKD"
5042. 5911
/gene="TM1748"
5042. 5911
/gene="TM1748"
/note="similar to GB:M57689 SP:P24139 GB:X56347 PID:143606
PID:40007 percent identity: 64.69; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36813.1"
/db_xref="GI:4982326"
/translation="MLRKNKTESIGWKAFLRFKKNKMAVIGGVFLIATLALAP
VTAPVPDEPHIRAFDESPKDTIFGDACGRDLFSITLSRNACITIGSGSEVYVI
IGGIGAVAGEFRGGMIDKFIWIVDIMEAFETFEVNLVYALGRGIFLFIAGLWG
NAGNARLIVGGVLYLKNSEFVGAACAASFTYIRKHILPNNMIGPLVNLASGVPA
NMTESGLAIVIGMGVPRPMPMSWGNLIGEGIMMAFPHLLIFPAVTFATLTISFTLAD
GIKDAFNRSRST"
5908. 6882
/gene="TM1749"
5908. 6882
/gene="TM1749"
/note="similar to SP:P04285 GB:X05491 GB:X52093 PID:47805
percent identity: 69.65; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36814.1"


```
/db_xref="GI:4982327"
/translation="MMELINYNLNKLVFHFVEIVKAVDYSYKLNKSGSLIGVSESG
SGKSVSLIRLTINRNGIYNDGEALIFPKDLKLNKEPRLIRKDSITIFOMNMTS
LNPIIRVGIQWNEPIIWHRLKMEAREPAITELLEVGPSPKPLNLPFOGSGMR
QRMVIMALACHPKLLIADDEPTALDVTLOAOIMELDLCKEYSMSYVETIHDLSVA
TNECRITITMYAGKIVEAPVEIKTLPHPYTKTLSTLSEIGRSGKGLVPIPNP
NPTKHPSCCKFHFRCSPAMEICQREEPVLNISENHRVACHLKGESK"
6852..7838
/gene="TM1750"
/length=1166
/seq_start=1
/seq_end=1166
/feature="CDS"
/feature_start=1
/feature_end=1166
/feature_note="similar to SP:P42065 PID:677944 GB:AL009126 percent identity: 76.43; identified by sequence similarity; putative"
/codon_start=1
/codon_end=1166
/translation="MPPOGGGIMKRPILQTVLKKYFPGKRLKAVDGSISTEIKGE
TLGIYSGSCCKSTIGRTILKLRPDGKIFREGDITNLNKEKPKPKKQIIFOD
PLGSLNPQWTVGRITIEDPLIHKISTKERRKVELLDVWIGREIFSEFHESSG
QOORIGIARALALNPKFIVCEPVALDVSIQAOIIDLEEQKMGISYLFIAHNL
VEHISHYVAVYLGKIVEYGDVYKIFLNPPIHPYRALIKSVKIPMDGOKRQFSLK
GELPSPIDLPKRCFRQTRCTEKKAICFEKEPELVEKHNHVSCHLVRSYR"
7898..8851
/gene="TM1751"
/length=953
/seq_start=1
/seq_end=953
/feature="CDS"
/feature_start=1
/feature_end=953
/feature_note="similar to SP:P23340 percent identity: 52.35; identified by sequence similarity; putative"
/codon_start=1
/codon_end=953
/translation="MGVDPFERNKILRGINIGALNLEAPNEGDMGVIRKDEPDIKE
AGFSHRIPRIMSTHAYAPPYKIMDRFRRVDEVINGLKGALVAVINIHYEELM
DPEKHFERIALMKQIADRYKDPETLEFELINETHNGULTPEKKNELLLEAKYRSI
DKKHTIITIGTAWGSIISALEKISVPRKEKNSIVTTHYINPEFTIQQEWGSGKWL
GKRWGSPDDQKHLIEFNPIEWSKRNKRPITIGFGAVRKADLESRLKWSFVVRKM
EKRRWSMAVWECGSGFVYDILRTKTNKDLLEALIGDSIE"
8844..9833
/gene="TM1752"
/length=989
/seq_start=1
/seq_end=989
/feature="CDS"
/feature_start=1
/feature_end=989
/feature_note="similar to SP:P23340 percent identity: 51.51; identified by sequence similarity; putative"
/codon_start=1
/codon_end=989
/translation="NNNTIIPWRGFNLLEAFSIKSTGNFKEDFLMMAQWDFNVIRIP
MCHLMSDRGNFPIREDFEKIDRVIFWGEKYGIHICISLHARAGSYNKEVEKTN
LWKDETQAEAFIHMSFIARRYKGISSTHLSNLIINPEPEPQIMASVEDHNSLIKRT
ITERRKIDPERLIIDIGYGNIPYVDLITENTVOSCGYIFSTVTHKAEVDSKDF
```

Query Match 94.7%; Score 903; DB 1; Length 10244;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 953; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGGGTTGTTGATCCCTTTGAAGAACAATAATTGGGAAGGCGCTTAATAGAAAT 60
Db 7898 ATGGGTGTTGATCCTTTGAAGAACAATAATTGGGAAGGCGCTTAATAGAAAT 7957
QY 61 GCGCTTGAAGCAACAATGAGGAGAGACTGGGAGTGTGATATAAAGATAGTCTTCGAC 120
Db 7958 GCGCTTGAAGCAACAATGAGGAGAGACTGGGAGTGTGATATAAAGATAGTCTTCGAC 8017
QY 121 ATTATTAAGAAGCGGTTTCTCTCATGTTCCAAATTCATAAGATGAGTACGACGCT 180
Db 121 ATTATTAAGAAGCGGTTTCTCTCATGTTCCAAATTCATAAGATGAGTACGACGCT 180
```

```
Db 8018 ATTATTAAGAAGCGGTTTCTCTCATGTTCCAAATTCATAAGATGAGTACGACGCT 8077
QY 181 TAGCGCTTCTCTCTTATTAATATCATGATCGCTTCTTCAAAAGAGTGAAGTATATA 240
Db 8078 TAGCGCTTCTCTCTTATTAATATCATGATCGCTTCTTCAAAAGAGTGAAGTATATA 8137
QY 241 AAGGAGCCCTGAAGAAGAGAGACTGGCTGTGTATATAATATTCATCCTCGAGAGATTA 300
Db 8138 AAGGAGCCCTGAAGAAGAGAGACTGGCTGTGTATATAATATTCATCCTCGAGAGATTA 8197
QY 301 ATGAATGATCCAGAAAGACACAAAGAAAGATTTCTGCTCTTGGAAACAATTTCTGAT 360
Db 8198 ATGAATGATCCAGAAAGACACAAAGAAAGATTTCTGCTCTTGGAAACAATTTCTGAT 8257
QY 361 CGTTATTAAGACATTCCTCCGAACTCTATTTTGGAAATTTGATGATACCTCAGGAAAT 420
Db 8258 CGTTATTAAGACATTCCTCCGAACTCTATTTTGGAAATTTGATGATACCTCAGGAAAT 8317
QY 421 CTACTCTCGGAAAGAAATGGAATGAACTCTTGAAGAGCTCTAAAGCTTATAGATCAATT 480
Db 8318 CTACTCTCGGAAAGAAATGGAATGAACTCTTGAAGAGCTCTAAAGCTTATAGATCAATT 8377
QY 481 GACAAAAGCACACTATATTAATAGGCACAGCTGAATGGGGGGTATATCTGCCCTTGA 540
Db 8378 GACAAAAGCACACTATATTAATAGGCACAGCTGAATGGGGGGTATATCTGCCCTTGA 8437
QY 541 AACTGCTCTCTCCCAAAATGGAAGAAATTTCTATAGTTTCAATTCACTACTCAATCT 600
Db 8438 AACTGCTCTCTCCCAAAATGGAAGAAATTTCTATAGTTTCAATTCACTACTCAATCT 8497
QY 601 TTGCAATTTACCATCAAGAGAGCTGAGTGGTGAAGAGATCGAGAATGTTGGGAGAGA 660
Db 8498 TTGCAATTTACCATCAAGAGAGCTGAGTGGTGAAGAGATCGAGAATGTTGGGAGAGA 8557
QY 661 AACTGGGAGATCTCCAGATGATGACAAAACATTTGATAGAGAAATTTATAGAGAA 720
Db 8558 AACTGGGAGATCTCCAGATGATGACAAAACATTTGATAGAGAAATTTATAGAGAA 8617
QY 721 TGCTCAAAAAGAAAGCAAAAGCAAAATTTACATAGAGAGTGGTGGCTACAGAAAAGCT 780
Db 8618 TGCTCAAAAAGAAAGCAAAAGCAAAATTTACATAGAGAGTGGTGGCTACAGAAAAGCT 8677
QY 781 GACCTTGAATCAAGATTAATAATGACCTCTTGTGCTGCGCAAAATGAGAAAAGAGAGA 840
Db 8678 GACCTTGAATCAAGATTAATAATGACCTCTTGTGCTGCGCAAAATGAGAAAAGAGAGA 8737
QY 841 TGGAGCTGGGCACTAGTGGCAATTTGTCCGTTTGGTGTATATGATCTGAGAGAAA 900
Db 8738 TGGAGCTGGGCACTAGTGGCAATTTGTCCGTTTGGTGTATATGATCTGAGAGAAA 8797
QY 901 ACCTGGAATTAAGATCTTTAGAGAGCTTTATAGAGAGAGATGACATTAATTA 954
Db 8798 ACCTGGAATTAAGATCTTTAGAGAGCTTTATAGAGAGAGATGACATTAATTA 8851
```

RESULT 5 AC079067 150882 bp DNA linear HTG 18-NOV-2000

LOCUS AC079067 Homo sapiens chromosome 4 clone RP11-77P11 map 4, WORKING DRAFT

DEFINITION AC079067.2 GI:11225393 SEQUENCE, 20 unordered pieces.

ACCESSION AC079067.2 HTG: HTGS_PHASE1; HTGS_DRAFT.

VERSION AC079067.2 HTG: HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS Homo sapiens.

SOURCE Homo sapiens.

ORGANISM Homo sapiens.

REFERENCE 1 (bases 1 to 150882) Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., JOURNAL Unpublished 2 (bases 1 to 150882) Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

TITLE
JOURNAL
COMMENT

Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavskiy, L.,
Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castile, A.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeTrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Ferreira, P.,
Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,
Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L.,
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Lahocque, K.,
Lamard, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,
Macdonald, P., Margus, N., McCarthy, M., McEwan, P., McKernan, K.,
McHeeters, R., Meldrum, J., Meneus, L., Mhova, T., Mlenga, V.,
Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sougnaz, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Travers, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A.,
Zimmer, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
Zimmer, A. and Zody, M.

Submitted (17-AUG-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 18, 2000 this sequence version replaced gi:9837996.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L10624

Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 137969 bases at least Q40
Consensus quality: 144918 bases at least Q30
Consensus quality: 147617 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 148982; sum-of-contigs
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 20 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 12294: contig of 12294 bp in length
* 12295 12394: gap of 100 bp
* 12395 14197: contig of 1803 bp in length
* 14198 14297: gap of 100 bp
* 14298 43602: contig of 29305 bp in length
* 43603 43702: gap of 100 bp
* 43703 44846: contig of 1144 bp in length
* 44847 44946: gap of 100 bp
* 44947 46353: contig of 1407 bp in length
* 46354 46453: gap of 100 bp
* 46454 48052: contig of 1599 bp in length
* 48053 48152: gap of 100 bp
* 48153 51116: contig of 2964 bp in length
* 51117 51216: gap of 100 bp
* 51217 52873: contig of 1657 bp in length
* 52874 52973: gap of 100 bp
* 52974 56008: contig of 3035 bp in length
* 56009 56108: gap of 100 bp
* 56109 59527: contig of 3419 bp in length

FEATURES
source

* 59528 59627: gap of 100 bp
* 59628 62795: contig of 3168 bp in length
* 62796 62895: gap of 100 bp
* 62896 67332: contig of 4437 bp in length
* 67333 67432: gap of 100 bp
* 67433 72873: contig of 5441 bp in length
* 72874 81642: contig of 8669 bp in length
* 81643 81742: gap of 100 bp
* 81743 87826: contig of 6084 bp in length
* 87827 87926: gap of 100 bp
* 87927 98128: contig of 10202 bp in length
* 98129 98228: gap of 100 bp
* 98229 109128: contig of 10900 bp in length
* 109129 109228: gap of 100 bp
* 109228 125996: contig of 16768 bp in length
* 125997 126096: gap of 100 bp
* 126097 149933: contig of 23837 bp in length
* 149934 150033: gap of 100 bp
* 150034 150882: contig of 849 bp in length.

FEATURES

Location/Qualifiers
1. 150882

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"

/clone="Rp11-77p11"
/clone_id="RPC1-11 Human Male BAC"
1. 12294

/note="assembly-fragment
clone_end:596
vector_side:left"

misc_feature
/note="assembly-fragment"
12395..14197

misc_feature
/note="assembly-fragment"
14298..43602

misc_feature
/note="assembly-fragment"
43703..44846

misc_feature
/note="assembly-fragment"
44947..46353

misc_feature
/note="assembly-fragment"
46454..48052

misc_feature
/note="assembly-fragment"
48153..51116

misc_feature
/note="assembly-fragment"
51217..52873

misc_feature
/note="assembly-fragment"
52974..56008

misc_feature
/note="assembly-fragment"
56109..59527

misc_feature
/note="assembly-fragment"
59628..62795

misc_feature
/note="assembly-fragment"
62896..67332

misc_feature
/note="assembly-fragment"
67433..72873

misc_feature
/note="assembly-fragment"
72974..81642

misc_feature
/note="assembly-fragment"
81743..87826

misc_feature
/note="assembly-fragment"
87927..98128

misc_feature
/note="assembly-fragment"
98229..109128

misc_feature
/note="assembly-fragment"
109229..125996

misc_feature
/note="assembly-fragment"
126097..149933

misc_feature
/note="assembly-fragment"
150034..150882

misc_feature
/note="assembly-fragment
vector_end:77
vector_side:right"

BASE COUNT

46319 a 27306 c 26924 g 48428 t 1905 others

ORIGIN

Query Match 2.4%; Score 23; DB 2; Length 150882;
 Best Local Similarity 100.0%; Pred. No. 0.4;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 563 AAAAAATCTATAGTTACATT 585
 ||||||||||||||||||
 Db 5557 AAAAAATCTATAGTTACATT 5579

RESULT 6

AC092659 160529 bp DNA linear PRI 01-MAR-2002
 LOCUS Homo sapiens BAC clone RP11-495115 from 4, complete sequence.
 DEFINITION AC092659 AC027044
 AC092659.2 GI:16303561
 VERSION HTG.
 KEYWORDS Homo sapiens.
 SOURCE Homo sapiens.
 ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074

REFERENCE
 AUTHORS Martin, S., Haakenson, W., and Elliott, G.
 TITLE The sequence of Homo sapiens BAC clone RP11-495115
 JOURNAL Unpublished (2001)
 REFERENCE
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (19-JUL-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (20-OCT-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (23-OCT-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Oct 20, 2001 this sequence version replaced gi:14916243.
 Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: saplens@wustl.edu
 Summary Statistics
 Center project name: H_NH0495115
 Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RP11-495115 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pleter de Jong and coworkers at <http://www.chori.org>
 VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-721f3; the clone sequenced to the right is AC079067. Actual start of this clone is at base position 1 of RP11-495115; actual end is at base position 160529 of RP11-495115.

The regions between bases 1 to 110 and 659 to 706 are represented only by PCR products of clone DNA. The region between 160353 to 160424 is covered only by a single plasmid subclone.

The sequence of AC027044 has been incorporated into AC092659.

FEATURES

source	Location/Qualifiers
1.160529	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-495115"
	/clone_1lb="RP11-11"
repeat_region	575..636
	/rpt_family="ERYL"
repeat_region	657..707
	/rpt_family="(GA)n"
repeat_region	701..1179
	/rpt_family="ERYL"
repeat_region	2817..2856
	/rpt_family="L2"
repeat_region	3473..3786
	/rpt_family="MER1_type"
repeat_region	3595..3650
	/rpt_family="AT_rich"
repeat_region	4674..5020
	/rpt_family="Alu"
repeat_region	4769..4814
	/rpt_family="A-rich"
repeat_region	6058..6353
	/rpt_family="ERYL"
repeat_region	6731..6773
	/rpt_family="AT_rich"
repeat_region	6834..6859
	/rpt_family="AT_rich"
repeat_region	7045..7078
	/rpt_family="AT_rich"
repeat_region	7795..7848
	/rpt_family="AT_rich"
repeat_region	7942..7972
	/rpt_family="AT_rich"
repeat_region	8306..8333
	/rpt_family="AT_rich"
repeat_region	8605..8638
	/rpt_family="(TTTTC)n"

repeat_region 8917. .8941
/rpt_family="(T)n"
repeat_region 10113. .10421
/rpt_family="Alu"
repeat_region 10398. .10421
/rpt_family="(A)n"
repeat_region 10448. .10481
/rpt_family="AT-rich"
repeat_region 11924. .11959
/rpt_family="(TAAA)n"
repeat_region 12981. .13071
/rpt_family="MIR"
repeat_region 13256. .13326
/rpt_family="(TA)n"
repeat_region 13451. .13484
/rpt_family="MIR"
repeat_region 14420. .14543
/rpt_family="L1"
repeat_region 14615. .14725
/rpt_family="L1"
repeat_region 14808. .14913
/rpt_family="MIR"
repeat_region 15012. .15040
/rpt_family="AT-rich"
repeat_region 15031. .15067
/rpt_family="L1"
repeat_region 15110. .15201
/rpt_family="L2"
repeat_region 15849. .15927
/rpt_family="L2"
repeat_region 18299. .18599
/rpt_family="Alu"
repeat_region 18821. .18859
/rpt_family="AT-rich"
repeat_region 18979. .19430
/rpt_family="L2"
repeat_region 19489. .19620
/rpt_family="(CAT)n"
repeat_region 19659. .19717
/rpt_family="(CAT)n"
repeat_region 21247. .21291
/rpt_family="L2"
repeat_region 21468. .21624
/rpt_family="L1"
repeat_region 21665. .22429
/rpt_family="L1"
repeat_region 22430. .22740
/rpt_family="Alu"
repeat_region 22712. .22757
/rpt_family="A-rich"
repeat_region 22741. .24033
/rpt_family="L1"
repeat_region 24034. .24319
/rpt_family="L1"
repeat_region 24320. .25563
/rpt_family="L1"
repeat_region 24988. .25014
/rpt_family="(TAAAA)n"
repeat_region 25576. .25988
/rpt_family="L1"
repeat_region 26091. .27707
/rpt_family="L1"
repeat_region 27719. .27838
/rpt_family="L1"
repeat_region 27839. .27873
/rpt_family="MALR"
repeat_region 27874. .28172

Query Match 2.4%; Score 23; DB 9; Length 160529;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 563 AAAAAATCTATAGTTACATT 585

Db 122669 AAAAAATCTATAGTTACATT 122691

RESULT 7

AC126611/c 178997 bp DNA linear HTG 07-JUL-2002
LOCUS Mus musculus chromosome UNK clone RP23-316N22, WORKING DRAFT
DEFINITION SEQUENCE, 10 unordered pieces.

ACCESSION

AC126611.1 GI:21702974
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

KEYWORDS

SOURCE

ORANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genome Center

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BA0316N22

Summary Statistics

Sequencing vector: M13: 0%
Sequencing vector: Plasmid: 100%
Chemistry: Dye-terminator ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 175866 bases at least Q40
Consensus quality: 176355 bases at least Q30
Consensus quality: 176586 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1844: contig of 1844 bp in length
* 1845 1944: gap of unknown length
* 1945 3071: contig of 1127 bp in length
* 3072 3171: gap of unknown length
* 3172 5257: contig of 2086 bp in length
* 5258 5357: gap of unknown length
* 5358 9354: contig of 3997 bp in length
* 9355 9454: gap of unknown length
* 9455 13669: contig of 4215 bp in length
* 13670 13769: gap of unknown length
* 13770 26976: contig of 13207 bp in length
* 26977 27076: gap of unknown length
* 27077 52620: contig of 25544 bp in length
* 52621 52720: gap of unknown length
* 52721 78960: contig of 26240 bp in length
* 78961 79060: gap of unknown length
* 79061 114898: contig of 35838 bp in length
* 114899 114998: gap of unknown length
* 114999 178997: contig of 63999 bp in length.
Location/Qualifiers
1. 178997

FEATURES

source

```

/misc_feature /organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-316N22"
1.1844
/misc_feature /note="assembly_name:Contig11"
1945.3071
/misc_feature /note="assembly_name:Contig12"
3172.5257
/misc_feature /note="assembly_name:Contig13"
5358.9354
/misc_feature /note="assembly_name:Contig14"
9455.13669
/misc_feature /note="assembly_name:Contig15"
13770.26976
/misc_feature /note="assembly_name:Contig16"
27077.52620
/misc_feature /note="assembly_name:Contig17"
52721.78960
/misc_feature /note="assembly_name:Contig18"
79061.114898
/misc_feature /note="assembly_name:Contig19"
114999.178997
BASE COUNT 53895 a 34886 c 33994 g 55211 t 1011 others
ORIGIN

```

```

Query Match 2.4% Score 23; DB 2; Length 178997;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 719 AATGTCACAAAGAACACAAAGA 741
|||||
Db 90950 AATGTCACAAAGAACACAAAGA 90928

```

```

RESULT 8 AC108661 194378 bp DNA linear HTG 13-JUL-2002
LOCUS AC108661
DEFINITION Rattus norvegicus clone CH230-231H7, *** SEQUENCING IN PROGRESS
ACCESSION AC108661
VERSION AC108661.3 GI:21737936
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE 1 (bases 1 to 194378)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alpbrechts,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbieri,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowler,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Burton,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
Cannon,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabriel,J.H., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Hornig,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korwar,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Louised,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 194378)
Unpublished
Direct Submission
Morley,K.C.
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 194378)
Morley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18846857.
--- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPZL
Center clone name: CH230-231H7
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 124895 bases at least Q40
Consensus quality: 129642 bases at least Q30
Consensus quality: 133710 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 79 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1598: contig of 1598 bp in length
1599: gap of unknown length
1699: contig of 1505 bp in length
3203: gap of unknown length
3303: gap of unknown length
3304: contig of 1198 bp in length
4501: gap of unknown length
4602: contig of 1131 bp in length
5733: gap of unknown length
5832: gap of unknown length
7389: contig of 1457 bp in length
7290: gap of unknown length
7390: contig of 1020 bp in length
8410: gap of unknown length
8510: contig of 1082 bp in length
9592: gap of unknown length
11185: contig of 1494 bp in length

```

```

* 11186 11285: gap of unknown length
* 11286 12518: contig of 1233 bp in length
* 12519 12618: gap of unknown length
* 12619 13822: contig of 1204 bp in length
* 13823 13922: gap of unknown length
* 13923 15081: contig of 1159 bp in length
* 15082 15181: gap of unknown length
* 15182 16462: contig of 1281 bp in length
* 16463 16562: gap of unknown length
* 16563 17779: contig of 1117 bp in length
* 17780 17779: gap of unknown length
* 17780 19196: contig of 1417 bp in length
* 19197 19296: gap of unknown length
* 19297 20635: contig of 1339 bp in length
* 20636 20735: gap of unknown length
* 20736 21962: contig of 1227 bp in length
* 21963 22062: gap of unknown length
* 22063 23138: contig of 1076 bp in length
* 23139 23238: gap of unknown length
* 23239 24790: contig of 1552 bp in length
* 24791 24890: gap of unknown length
* 24891 26080: contig of 1190 bp in length
* 26081 26180: gap of unknown length
* 26181 27612: contig of 1432 bp in length
* 27613 27712: gap of unknown length
* 27713 28829: contig of 1117 bp in length
* 28830 28929: gap of unknown length
* 28930 30142: contig of 1213 bp in length
* 30143 30242: gap of unknown length
* 30243 31838: contig of 1596 bp in length
* 31839 31938: gap of unknown length
* 31939 33398: contig of 1460 bp in length
* 33399 33498: gap of unknown length
* 33499 34808: contig of 1310 bp in length
* 34809 34908: gap of unknown length
* 34909 36063: contig of 1155 bp in length
* 36064 36163: gap of unknown length
* 36164 37521: contig of 1358 bp in length
* 37522 37621: gap of unknown length
* 37622 39071: contig of 1450 bp in length
* 39072 39171: gap of unknown length
* 39172 40179: contig of 1008 bp in length
* 40180 40279: gap of unknown length
* 40280 42264: contig of 1985 bp in length
* 42265 42364: gap of unknown length
* 42365 43904: contig of 1540 bp in length
* 43905 44004: gap of unknown length
* 44005 46127: contig of 2123 bp in length
* 46128 46227: gap of unknown length
* 46228 48010: contig of 1783 bp in length
* 48011 48110: gap of unknown length
* 48111 49843: contig of 1733 bp in length
* 49844 49943: gap of unknown length
* 49944 52363: contig of 2420 bp in length
* 52364 52463: gap of unknown length
* 52464 54700: contig of 2237 bp in length
* 54701 54800: gap of unknown length
* 54801 56327: contig of 1527 bp in length
* 56328 56427: gap of unknown length
* 56428 57929: contig of 1502 bp in length
* 57930 58029: gap of unknown length
* 58030 59377: contig of 1348 bp in length
* 59378 59477: gap of unknown length
* 59478 61920: contig of 2443 bp in length
* 61921 62020: gap of unknown length
* 62021 63992: contig of 1972 bp in length
* 63993 64092: gap of unknown length
* 64093 66233: contig of 2141 bp in length
* 66234 66333: gap of unknown length
* 66334 68509: contig of 2176 bp in length
* 68510 68609: gap of unknown length
* 68610 70025: contig of 1416 bp in length
* 70026 70125: gap of unknown length

```

```

* 70126 71550: contig of 1425 bp in length
* 71551 71650: gap of unknown length
* 71651 72860: contig of 1210 bp in length
* 72861 72960: gap of unknown length
* 72961 74649: contig of 1689 bp in length
* 74650 74749: gap of unknown length
* 74750 77133: contig of 2384 bp in length
* 77134 77233: gap of unknown length
* 77234 78973: contig of 1740 bp in length
* 78974 79073: gap of unknown length
* 79074 81411: contig of 2338 bp in length
* 81412 81511: gap of unknown length
* 81512 83705: contig of 2194 bp in length
* 83706 83805: gap of unknown length
* 83806 86534: contig of 2729 bp in length
* 86535 86635: gap of unknown length
* 86635 88506: contig of 1872 bp in length
* 88507 88607: gap of unknown length
* 88607 89903: contig of 1297 bp in length

Query Match          2.3%: Score 22; DB 2; Length 194378;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 900 AACCTGCAATAAGATCTTTA 921
    |||||||
Db 82561 AACCTGCAATAAGATCTTTA 82582

RESULT 9
LOCUS AR077713 51 bp DNA linear PAT 31-AUG-2000
DEFINITION Sequence 3 from patent US 5962258.
ACCESSION AR077713
VERSION AR077713.1 GI:10004459
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
  1 (bases 1 to 51)
AUTHORS Mathur,E.J. and Lam,D.E.
TITLE Carboxymethyl cellulase from thermotoga maritima
JOURNAL Patent: US 5962258-A 3 05-OCT-1999;
FEATURES
  source
    1..51
    /organism="Unknown"
BASE COUNT 14 a 6 c 14 g 17 t
ORIGIN

Query Match          2.2%: Score 21; DB 6; Length 51;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGGTGATCCTTTGAA 21
    |||||||
Db 31 ATGGGTGATCCTTTGAA 51

RESULT 10
LOCUS AR096553 51 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 3 from patent US 6008032.
ACCESSION AR096553
VERSION AR096553.1 GI:10025459
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE
  1 (bases 1 to 51)
AUTHORS Mathur,E.J. and Lam,D.E.
TITLE Carboxymethyl cellulase from Thermotoga maritima
JOURNAL Patent: US 6008032-A 3 28-DEC-1999;
FEATURES
  Location/Qualifiers

```

source	1..51	/organism="unknown"							
BASE COUNT	14 a	6 c	14 g	17 t					
ORIGIN									
Query Match	2.2%;	Score 21;	DB 6;	Length 51;					
Best Local Similarity	100.0%;	Pred. No. 7.6;							
Matches 21;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
OY	1	ATGGGTGTTGATCCYTTTGAA 21							
Db	31	ATGGGTGTTGATCCYTTTGAA 51							
RESULT 11									
LOCUS	ARI57662		51 bp	DNA	Linear				
DEFINITION	Sequence 3 from patent US 6245547.								
ACCESSION	ARI57662								
VERSION	ARI57662.1	GI:16218625							
KEYWORDS									
SOURCE	Unknown.								
ORGANISM	Unknown.								
REFERENCE	Unclassified.								
AUTHORS	1 (bases 1 to 51)								
TITLE	Mathur,E.J. and Lam,D.E.								
JOURNAL	Carboxymethyl cellulose from thermotoga maritima								
FEATURES	Patent: US 6245547-A 3 12-JUN-2001;								
source	Location/Qualifiers								
	1..51								
	/organism="unknown"								
BASE COUNT	14 a	6 c	14 g	17 t					
ORIGIN									
Query Match	2.2%;	Score 21;	DB 6;	Length 51;					
Best Local Similarity	100.0%;	Pred. No. 7.6;							
Matches 21;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
OY	1	ATGGGTGTTGATCCYTTTGAA 21							
Db	31	ATGGGTGTTGATCCYTTTGAA 51							
RESULT 12									
LOCUS	HS86F14		106571 bp	DNA	linear				
DEFINITION	Human DNA sequence from PAC 86F14 on chromosome 1q23-1q24. Contains								
ACCESSION	coagulation factor V, ESTs and STS.								
VERSION	Z99572								
KEYWORDS	Z99572.1 GI:2769646								
SOURCE	1q23-1q24; Blood coagulation factor; factor V.								
ORGANISM	Homo sapiens.								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.								
TITLE	1 (bases 1 to 106571)								
JOURNAL	Bird.C.								
COMMENT	Direct Submission								
	submitted (13-JAN-1998) Chromosome 1 Project Group								
	(http://www.sanger.ac.uk/HGP/Chr1/) Sanger Centre, Hinxton,								
	Cambridgeshire, CB10 1SA, UK. E-mail enquiries:								
	humangenes@sanger.ac.uk; UK Clone requests: clonequest@sanger.ac.uk								
	On Jan 13, 1998 this sequence version replaced gi:2578147.								
	IMPORTANT: This sequence is not the entire insert of clone 86F14.								
	It may be shorter because we only sequence overlapping sections								
	once, or longer because we arrange for a small overlap between								
	neighbouring submissions.								
	During sequence assembly data is compared from overlapping clones.								
	Where differences are found these are annotated as variations								
	together with a note of the overlapping clone name. Note that the								
	variations annotated may not be found in the sequence submission								
	corresponding to the overlapping clone as we submit sequences with								
	only a small overlap as described above.								

```

This sequence was generated from part of bacterial clone contigs of human chromosome 1 constructed by the Sanger Centre chromosome 1 mapping group. Further information can be found at http://www.sanger.ac.uk/HGP/chr1/
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The true right end of clone 206D15 is at 104.
The true right end of clone 86F14 is at 106571.
86F14 is from the library RPCI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong.
For further details see http://bacpac.med.buffalo.edu/.
Location/Qualifiers
1..106571
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q23-1q24"
/clone="RP1-86F14"
/clone_id="RPCI-1"
810..1090
/note="AluX repeat: matches 1..302 of consensus"
1270..1360
/note="L1 repeat: matches 1..84 of consensus"
1410..1587
/note="L1 repeat: matches 139..323 of consensus"
1575..1858
/note="L1 repeat: matches 265..568 of consensus"
1876..2745
/note="L1 repeat: matches 921..1 of consensus"
2597..2793
/note="L1 repeat: matches 5390..5194 of consensus"
2854..3352
/note="L1 repeat: matches 687..170 of consensus"
2971..3352
/note="L1 repeat: matches 570..170 of consensus"
3327..3359
/note="L1 repeat: matches 1..218 of consensus"
3580..3873
/note="L1 repeat: matches 1..299 of consensus"
3877..4060
/note="L1 repeat: matches 193..10 of consensus"
6468..6575
/note="L1 repeat: matches 22..145 of consensus"
7024..7238
/note="L1 repeat: matches 84..299 of consensus; incomplete repeat"
7834..8126
/note="L1 repeat: matches 2..295 of consensus"
10054..10223
/note="L1 repeat: matches 89..262 of consensus"
11725..12081
/note="L1 repeat: matches 371..1 of consensus"
12196..12497
/note="L1 repeat: matches 303..1 of consensus"
13253..13367
/note="L1 repeat: matches 4..127 of consensus"
13398..15749
/note="L1 repeat: matches 2577..4940 of consensus"
16076..16554
/note="L1 repeat: matches 4910..5390 of consensus"
16409..17294
/note="L1 repeat: matches 1..891 of consensus"
17808..19825
/note="L1 repeat: matches 3338..5390 of consensus"
19683..20571
/note="L1 repeat: matches 1..890 of consensus"
22732..22827

```

```
/note="LIMB4 repeat: matches 797. .902 of consensus"
22748. .22851
/note="LIMB5 repeat: matches 812. .922 of consensus"
<25068. .25343
prim_transcript
/note="match: H61071 H69565"
25302. .25381
repeat_region
/note="MIR repeat: matches 49. .131 of consensus"
25527. .25703
repeat_region
/note="MER45 repeat: matches 1. .178 of consensus"
25932. .26062
repeat_region
/note="MIR repeat: matches 18. .138 of consensus"
26055. .26264
repeat_region
/note="L1 repeat: matches 2971. .2759 of consensus"
26266. .26378
repeat_region
/note="LIM47 repeat: matches 1017. .895 of consensus"
27046. .27579
prim_transcript
/note="match: multiple ESTs; match: R71060 H69028 R82016
R10102 R82280 R82281; match: R82066 H69792 H79486 AA506861
T78091; match: W03874 H74282 DB5329"
278091. .27851
CDS
complement(join(27431. .27577,28562. .28744,31530. .31681,
33638. .33782,36315. .36470,36919. .37022,37955. .38026,
39019. .39135,41033. .41212,42726. .42936,43904. .44140,
46624. .46798,53412. .56232,57414. .57626,59560. .59710,
62919. .63133,63758. .63857,65675. .65852,68300. .68465,
69764. .69985,72271. .72414,73672. .73884,85339. .85461,
95549. .95640,99347. .99504))
/codon_start=1
/product="factor V"
/protein_id="CAB16748.1"
/db_xref="GI:2769647"
/db_xref="SPTREMBL:O43737"
/translation="MPCGCPRIWLVVIGTGVNGVSGSGSTEAOLRQFVVAAGISMS
YRPEPTNSLNVTSFKIVREYEPFKKKKPOSTISGLGPIPLAVNGDITVHF
KKKAKPLSIHPQIGIRSKLSEGAYLDTFAEKMDVAAGREYITWMSISEBGF
TDDPPLTHIYSHENLIEDNSGLIKCKGTLEGQTKEDQIVLEAVE
DEKSSOSSILMYNGVNGTMDPITVCAIDSLIWMHLSGSGELSIHNGQVL
RONHKSVAITLVSATSTANNVGEQKIIISLPLKQAGMAYIDIKNCQKTR
NKKITREORRRHMKREYFIAAEVITWDAVPIAMQDVKIYFKNNASPSYIPGV
KMYVTEDESESTKRTVNNMKEDGILGRTIYKNIILEDEPTENDACILRRYSD
TSPYEDEVNSFTSGRNNMTIYAVPGETTYTKNIILEDEPTENDACILRRYSD
VIMDRASGLIGLILICKSRSLDRGIDIRADIEQAAFAVDEKSNYLEDNINKE
CENPDEVRKDDPEFYENIMSTINGVPSITTLGFCFDVQVMHCSVGTONEILTI
HFTGHSFIYGRHEDTLTLPYMGESVITVMDVGTWMLTSMNSSPRSKRLKRFEDY
KCTPDDEDSYEIEFPEPESTVATRKMDLEDEDEADYQNRKLAALGIRSEF
NSSLNDEEENFLTALALNGTGFVSNDIIVGNSPSNISKTYTNLLAEQAP
FHOQATYAGSPPLHLIGKSNVLSNSTAHSHPSPEDIDPDLVGTIRLLSAGE
FSQEHAKHKGPRVQDAKHFNMKLLAHVGHLSLSDTSPSCGMRPEDVPLMD
TSPSMRPMKDDPDLILLKQSNSSKILVGRWHLSEKSTETIDTDETRAVNMIL
ISPNASRAMGEESTPLANKRGQSGHPRVYRKLSEVLAHKNESLPLDLMLTLP
KKKEKTHHAPLSPRTPHPLRSEAYTESEERLKLSHLVAKNESLPLDLMLTLP
MDGWTASLIDHNNSSNDTGAQSPGILQYVPEHQTPEIOPDDHMSDPSH
RSSPPELSEMLEYDRSHKSPRDIISOMSPSEHVMQVTSIPOLSTLSPELSONL
SPDLSTLSPELIQRKLSPALQMPISPOLSTLSPELSHLSTLSDTSONLSPEL
SQNLSPALGOMPLSPDLSTLSPDLSTLSPDLSTLSPDLSTLSPDLSTLSPDL
ISPDLSHTLSPDLSTLSPDLSTLSPDLSTLSPDLSTLSPDLSTLSPDLSTL
LSQNLSTLSPDLSTLSPDLSTLSPDLSTLSPDLSTLSPDLSTLSPDLSTL
TSPDLSTLSTLSPDLSTLSPDLSTLSPDLSTLSPDLSTLSPDLSTLSPDLST
SPDLSTLSTLSPDLSTLSPDLSTLSPDLSTLSPDLSTLSPDLSTLSPDLST
DVRNTINSSRDPNIAAMYLRNNGNRRNYIAEELISVSEVQEDIEDIDIP
EDTYTKVYRKLYLSTFTTRDPRGEYEHGLGILPIIRAEVDVIOVRKNLASRY
SLAHGLSYERKSEGTEDSPMEVEDNAVONNSYTYVMAHTEPESGESGESACR
AMAYYSANPEKDIHSLGLPLIICQGLIHKSNMMDREYLLWLTEDERSKRY
EKKRSRMLTSEMKKSHFHAINGMIVLPGKIKEMKSGKMWLLNTEVENQRAQGP
HFGQITLNGENKOHQVWPLPGSEFTLEMAKSGKMWLLNTEVENQRAQGP
LHNDRCRMGLSTGIISDQIKASELIVYERLARLNNNGSVENKAAAEFA
SKPWIOVDMQEVITITIGOGAKHYIKSCYTFEYVAYASNNQIMQIFGNSRNM
YFNGNSDASTIKENQEPPIVARIIRISPRVAVRPILELOCEVENGSGTPLGMN
GKIENKQITASSFKKSWMGDYWEFARLNAQGVNAQKANNKQWLELDLKLK
ITAITOGCKSLSEMYKSYTIHYSEQGEWKYRIKSSVADKIFEGNNTKGHAVN
FENPITSRPIRIVPTKWNOSIALRLEFGDLY"
27629. .27788
repeat_region
/note="AluJo repeat: matches 302. .140 of consensus;
incomplete repeat"
```

```
repeat_region 28002. .28176
/note="MIR repeat: matches 75. .259 of consensus"
repeat_region 29041. .29084
/note="MIR2 repeat: matches 103. .146 of consensus"
repeat_region 31903. .31988
/note="MIR repeat: matches 104. .17 of consensus"
repeat_region 34738. .35036
/note="AluSc repeat: matches 299. .1 of consensus"
repeat_region 35051. .35542
/note="L1 repeat: matches 3215. .3729 of consensus"
repeat_region 36628. .36808
/note="MIR repeat: matches 40. .235 of consensus"
repeat_region 38333. .38451
/note="AluJo repeat: matches 1. .134 of consensus;
incomplete repeat"
repeat_region 40215. .40272
/note="MIR2 repeat: matches 1. .53 of consensus"
repeat_region 43274. .43327
/note="MIR2 repeat: matches 146. .93 of consensus"
repeat_region 43456. .43548
/note="LIMB8 repeat: matches 1038. .944 of consensus"
repeat_region 44951. .46774
/note="L1 repeat: matches 3613. .5390 of consensus"
repeat_region 46627. .46980
/note="LIM47 repeat: matches 2. .364 of consensus"
repeat_region 46983. .47446

Query Match 2.2%; Score 21; DB 9; Length 106571;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 786 TGAATCAAGATTAATAATGAC 806
Db 56614 TGAATCAAGATTAATAATGAC 56634
|||||
|||||

RESULT 13
AC007870/c 134757 bp DNA linear PRI 12-AUG-1999
LOCUS AC007870
DEFINITION Genomic sequence for Homo sapiens clone 4b6, complete sequence.
ACCESSION AC007870
VERSION AC007870.3 GI:5731403
KEYWORDS HTC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1. Haberman, K., Spiegel, L., de la Bastide, M., Dedhia, N.,
Rodriguez, M., Shah, R., Schutz, K., Shekher, M., Spiegel, L., Swaby, L.,
Vill, D. and McCombie, W. R.
Genomic Sequence for Homo sapiens clone 4b6
Unpublished
2. (bases 1 to 134757)
McCombie, W. R.
Direct Submission
Submitted (19-JUN-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
3. (bases 1 to 134757)
McCombie, W. R.
Direct Submission
Submitted (12-AUG-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
On Aug 12, 1999 this sequence version replaced gi:5306231.
From 107690-107780, the shotgun assembly of 4b6 was single-stranded
with single chemistry. Cloning vector was pBelOAc11.
location/Qualifiers
1. .134757
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
source
```


BASE COUNT 41775 a 23641 c 25490 g 43851 t
 ORIGIN /clone.lib="PCR-screenable Research Genetics Library"

Query Match 2.2%; Score 21; DB 9; Length 134757;
 Best Local Similarity 100.0%; Pred. No. 5.4;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TATTGGAGAGGCATTAAATA 52
 Db 130408 TATTGGAGAGGCATTAAATA 130388

RESULT 14
 AL355177 134781 bp DNA linear ROD 29-JUN-2002
 LOCUS Mouse DNA sequence from clone RP21-61H21 on chromosome 5, complete
 DEFINITION sequence.
 ACCESSION AL355177
 VERSION AL355177.14 GI:14529794
 KEYWORDS HTG.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerys@sanger.ac.uk
 On Jun 22, 2001 this sequence version replaced g1:11610998.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Info: Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 from the RPI-21 Mouse PAC Library constructed by the group of Pieter de Jong.
 For further details see http://www.chori.org/bacpac/home.htm
 VECTOR: pPAC4

----- Genome Center
 Center: UK Medical Research Council
 Center code: UK-MRC
 Web site: http://mrcseq.har.mrc.ac.uk
 Contact: mouse@har.mrc.ac.uk

FEATURES
 source
 1. 134781
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="5"
 /clone="RP21-61H21"
 /clone.lib="RPI-21"
 BASE COUNT 37151 a 29553 c 30055 g 38022 t
 ORIGIN

Query Match 2.2%; Score 21; DB 10; Length 134781;
 Best Local Similarity 100.0%; Pred. No. 5.4;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 387 ATTTTGAATCTGAATGA 407
 Db 35844 ATTTTGAATCTGAATGA 35824

RESULT 15
 AL513082 137154 bp DNA linear PRI 12-JUN-2001
 LOCUS Human DNA sequence from clone RP11-229N15 on chromosome Xq24-25, complete sequence.
 DEFINITION
 ACCESSION AL513082
 VERSION AL513082.10 GI:14422228
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Direct Submission
 Submitted (11-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerys@sanger.ac.uk
 On Jun 13, 2001 this sequence version replaced g1:14141422.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Info: Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX
 RP11-229N15 is from the library RPI-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-229N15 The true left end of clone RP11-161K18 is at 87431 in this sequence.

FEATURES
 source
 1. 137154
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="q24-25"
 /clone="RP11-229N15"
 /clone.lib="RPI-11.1"
 repeat_region 1. 1927
 /note="LIM1 repeat: matches 4363. 6304 of consensus"
 repeat_region 1961. 2667
 /note="LIM2 repeat: matches 5353. 6073 of consensus"
 repeat_region 2670. 3022
 /note="THE1B repeat: matches 1. 364 of consensus"
 repeat_region 3023. 3258
 /note="LIM2 repeat: matches 6071. 6307 of consensus"
 repeat_region 3282. 4586
 /note="LIM1 repeat: matches 4643. 5968 of consensus"
 misc_feature 4450. 4843

```
/note="match: GSS: Em:AQ116765"
4716..6342
repeat_region /note="L1M1 repeat: matches 2845..4472 of consensus"
misc_feature /note="match: STS: Em:G25694"
7020..7364
misc_feature /note="match: STS: Em:G11556 Em:G26337"
7179..7565
misc_feature /note="match: STS: Em:AQ054060"
complement(7212..7657)
7317..7636
misc_feature /note="match: GSS: Em:A2694362"
7650..8451
repeat_region /note="L1M1 repeat: matches 4463..5230 of consensus"
8452..8861
repeat_region /note="L1M1 repeat: matches 4..426 of consensus"
8862..9531
repeat_region /note="L1M1 repeat: matches 5230..5998 of consensus"
9586..9654
repeat_region /note="L1 repeat: matches 2262..2327 of consensus"
9655..9952
repeat_region /note="L1 repeat: matches 3..304 of consensus"
9953..10352
repeat_region /note="L1 repeat: matches 2327..2750 of consensus"
10508..10587
repeat_region /note="L1 repeat: matches 2 mer at 67% conserved"
10826..10972
repeat_region /note="L1 repeat: matches 31..178 of consensus"
11890..12256
repeat_region /note="L1 repeat: matches 1..365 of consensus"
14740..14799
repeat_region /note="L1 repeat: matches 1..365 of consensus"
14807..16822
repeat_region /note="L1 repeat: matches 4131..6146 of consensus"
16822..19172
repeat_region /note="L1 repeat: matches 1777..4135 of consensus"
21608..21672
repeat_region /note="L1 repeat: matches 86..148 of consensus"
22071..22375
repeat_region /note="L1 repeat: matches 1..305 of consensus"
22850..22877
repeat_region /note="L1 repeat: matches 2 mer tg 100% conserved"
23078..23131
repeat_region /note="L1 repeat: matches 2 mer ta 74% conserved"
23148..23523
repeat_region /note="L1 repeat: matches 1..364 of consensus"
23524..25074
repeat_region /note="L1 repeat: matches 1..1580 of consensus"
25075..25447
repeat_region /note="L1 repeat: matches 1..364 of consensus"
complement(25519..25960)
misc_feature /note="match: STS: Em:HSC685"
27581..27822
misc_feature /note="match: STS: Em:AQ877624"
27917..29879
repeat_region /note="L1 repeat: matches -1422..-1192 of consensus"
30125..30409
repeat_region /note="L1 repeat: matches 968..3008 of consensus"
30450..33568
repeat_region /note="L1 repeat: matches 1..285 of consensus"
34046..34050
repeat_region /note="L1 repeat: matches 3003..6156 of consensus"
This sequence represents the duplicated flanking sequence
of the IS2"
34565..35225
repeat_region /note="L1 repeat: matches 1628..2327 of consensus"
35262..35440
repeat_region /note="L1 repeat: matches 2515..2682 of consensus"
38611..38689
repeat_region /note="L1 repeat: matches 176..249 of consensus"
38715..38881
```

```
misc_feature /note="L1P13 repeat: matches -150..4 of consensus"
complement(38932..39495)
repeat_region /note="match: STS: Em:HSC67C4"
39223..40001
repeat_region /note="L1P13 repeat: matches -651..158 of consensus"
40235..41809
repeat_region /note="L1P12 repeat: matches 522..2208 of consensus"
41810..42214
repeat_region /note="L1 repeat: matches 1..426 of consensus"
42224..43684
repeat_region /note="L1 repeat: matches 198..1648 of consensus"
43685..44049
repeat_region /note="L1 repeat: matches 1..371 of consensus"
44050..44238
repeat_region /note="L1 repeat: matches 9..198 of consensus"
44265..44651
repeat_region /note="L1 repeat: matches 1..408 of consensus"
44652..46249
repeat_region /note="L1 repeat: matches 2204..3791 of consensus"
46252..46353
repeat_region /note="L1 repeat: matches 3794..3897 of consensus"
46351..46955
repeat_region /note="L1 repeat: matches 3816..4432 of consensus"
46924..47680
repeat_region /note="L1 repeat: matches 3961..4718 of consensus"
47686..48489
repeat_region /note="L1 repeat: matches 5302..6137 of consensus"
48518..50273
repeat_region /note="L1 repeat: matches 4395..6157 of consensus"
50967..51212
repeat_region /note="L1 repeat: matches 28..252 of consensus"
52867..52958
repeat_region /note="L1 repeat: matches 2 mer aa 67% conserved"
53142..53324
repeat_region /note="L1 repeat: matches 6..189 of consensus"
53486..53618
repeat_region /note="L1 repeat: matches 20..145 of consensus"
55100..55525
repeat_region /note="L1 repeat: matches 20..145 of consensus"
55105..55472
repeat_region /note="L1 repeat: matches 20..145 of consensus"
55327..55423
repeat_region /note="L1 repeat: matches 400..494 of consensus"
56269..56320
repeat_region /note="L1 repeat: matches 26 copies 2 mer ca 96% conserved"
57538..57957
repeat_region /note="L1 repeat: matches 1..418 of consensus"
57958..58214
repeat_region /note="L1 repeat: matches 307..566 of consensus"
58395..58472
repeat_region /note="L1 repeat: matches 30..101 of consensus"
59578..60160
repeat_region /note="L1 repeat: matches 102..662 of consensus"
60270..61693
repeat_region /note="L1 repeat: matches 673..2097 of consensus"
62055..62385
repeat_region /note="L1 repeat: matches 139..466 of consensus"
62386..62431
repeat_region /note="L1 repeat: matches 366..411 of consensus"
62435..62504
repeat_region /note="L1 repeat: matches 2 mer tg 80% conserved"
62529..62910
```

Query Match 2.2%; Score 21; DB 9; Length 137154;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 386 TATTTTGAATTCGATG 406
DB 11746 TATTTTGAATTCGATG 11766

```

RESULT 16
AC079562      140795 bp   DNA      linear   HTG 02-SEP-2000
LOCUS          Mus musculus clone Rp23-465P16, WORKING DRAFT SEQUENCE. 24
DEFINITION    Unordered pieces.
ACCESSION     AC079562
VERSION       AC079562.1   GI:9964927
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        Mus musculus.
ORGANISM      Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 140795)
AUTHORS       DOE Joint Genome Institute.
TITLE         Sequencing of Mouse
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 140795)
AUTHORS       DOE Joint Genome Institute.
TITLE         Direct Submission
JOURNAL       Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT       -----Genome Center
              Center: Joint Genome Institute
              Center Code: JGI
              Web site: http://www.jgi.doe.gov
              -----
              Project Information
              Center Project Name: 1915808
              Center clone name: RPCI-23_465P16
              -----
              Summary Statistics
              Consensus quality: 112243 bases at least Q40
              Consensus quality: 123963 bases at least Q30
              Consensus quality: 126961 bases at least Q20
              Estimated insert size: 187000; agarose-1p estimation
              Estimated insert size: 138495; sum-of-ctrls estimation
              Quality coverage: 6.23 in Q20 bases; agarose-1p estimation
              Quality coverage: 8.41 in Q20 bases; sum-of-ctrls estimation.
              NOTE: This is a 'working draft' sequence. It currently
              * consists of 24 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              -----
              1 1202: contig of 1202 bp in length
              * 1203 1302: gap of unknown length
              * 1303 2446: contig of 1144 bp in length
              * 2447 2546: gap of unknown length
              * 2547 3844: contig of 1298 bp in length
              * 3845 3944: gap of unknown length
              * 3945 5015: contig of 1071 bp in length
              * 5016 5115: gap of unknown length
              * 5116 6372: contig of 1257 bp in length
              * 6373 6472: gap of unknown length
              * 6473 7613: contig of 1141 bp in length
              * 7614 7713: gap of unknown length
              * 7714 8895: contig of 1182 bp in length
              * 8896 8995: gap of unknown length
              * 8996 10604: contig of 1609 bp in length
              * 10605 10704: gap of unknown length
              * 10705 11824: contig of 1120 bp in length
              * 11825 11924: gap of unknown length
              * 11925 13682: contig of 1758 bp in length
              * 13683 13782: gap of unknown length
              * 13783 16337: contig of 2555 bp in length
              * 16338 16437: gap of unknown length
              * 16438 18700: contig of 2263 bp in length
              * 18701 18800: gap of unknown length
              * 18801 24180: contig of 5380 bp in length
              * 24181 24280: gap of unknown length
              * 24281 30054: contig of 5774 bp in length

```

```

* 30055 30154: gap of unknown length
* 30155 36771: contig of 6617 bp in length
* 36772 36871: gap of unknown length
* 36872 41167: contig of 4296 bp in length
* 41168 41267: gap of unknown length
* 41268 48107: contig of 6839 bp in length
* 48107 48206: gap of unknown length
* 48206 54870: contig of 6664 bp in length
* 54870 54871: gap of unknown length
* 54871 61950: contig of 6980 bp in length
* 61951 62050: gap of unknown length
* 62051 72051: contig of 10001 bp in length
* 72052 72151: gap of unknown length
* 72152 86891: contig of 14740 bp in length
* 86892 86991: gap of unknown length
* 86992 103983: contig of 16992 bp in length
* 103984 104083: gap of unknown length
* 104084 123755: contig of 19672 bp in length
* 123756 123855: gap of unknown length
* 123856 140795: contig of 16940 bp in length.
Location/Qualifiers
1. 140795
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-465P16"
/clone_lib="RPCI mouse BAC library 23"
BASE COUNT 33775 a 36314 c 35670 g 32605 t 2431 others
ORIGIN
Query Match 2.2%; Score 21; DB 2; Length 140795;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 725 CAAAGAAAGACAAAGACCAA 745
Db 16702 CAAAGAAAGACAAAGACCAA 16722
RESULT 17
LOCUS          AC010534      146385 bp   DNA      linear   PRI 26-JAN-2001
DEFINITION    Homo sapiens chromosome 16 clone RP11-259E8, complete sequence.
ACCESSION     AC010534
VERSION       AC010534.7   GI:12545287
KEYWORDS      HTG.
SOURCE        Homo sapiens.
ORGANISM      Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE     1 (bases 1 to 146385)
AUTHORS       DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE         Direct Submission
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 146385)
AUTHORS       DOE Joint Genome Institute.
TITLE         Direct Submission
JOURNAL       Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT       3 (bases 1 to 146385)
              DOE Joint Genome Institute and Stanford Human Genome Center.
              Direct Submission
              Submitted (26-JAN-2001) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
              On Jan 26, 2001 this sequence version replaced gi:9954624.
              Draft Sequence Produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Finishing Completed at Stanford Human Genome Center
              www-shgc.stanford.edu
              Quality: Phrap Quality >=40 99.8% of Sequence;
              Estimated Total Number of Errors is 0.2.
              STS Content:
              SHGC-80918 G53965.
              Location/Qualifiers

```

```

source
1. .146385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-259E8"

BASE COUNT      45853 a 27125 c 26308 g 47099 t
ORIGIN

Query Match      2.2%; Score 21; DB 9; Length 146385;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      9 TGATCCTTTTGAAGGACAA 29
|||||
Db 129419 TGATCCTTTTGAAGGACAA 129439

RESULT 18
AC124046/c      166496 bp      DNA      linear      PRI 27-JUN-2002
LOCUS      Homo sapiens chromosome 3 clone RP11-352B8, complete sequence.
DEFINITION
AC124046
AC124046.2 GI:21617716
VERSION
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 166496)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.
and Haugen, E.D.
Direct Submission
Unpublished
2 (bases 1 to 166496)
Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
Direct Submission
Submitted (08-JUN-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 166496)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.
and Haugen, E.D.
Direct Submission
Submitted (27-JUN-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Jun 27, 2002 this sequence version replaced gi:21358747.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
COMMENT

Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Project Information
Center project name: chr-3
Center clone name: RP11-352B8 (bc0803)
Summary Statistics
Sequencing vector: plasmid; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 166409 bases at least Q40
Consensus quality: 166495 bases at least Q30
Consensus quality: 166496 bases at least Q20
Insert size: 166496; sum-of-contigs
Quality coverage: 10.7x in Q20 bases; sum-of-contigs

Overlapping Sequences:
5' : Mapping in progress
3' : RP11-281N1 (UWGC:bc0639) AC104644

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.

```

Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

NsII		BglII		EcoRI	
SeqDermap	Fgprpnt	SeqDermap	Fgprpnt	SeqDermap	Fgprpnt
26396	26807	2444	2476	8656	8843
18300	17543	2067	1977	6	<800
6322	6279	4828	4674	14125	14135
748	765	127	<800	1854	1846
180	<800	2220	2187	6813	6655
6666	6665	511	<800	229	<800
3924	4052	2289	2304	3929	3846
6109	6132	2144	2187	5263	5161
707	<800	4262	4237	4083	4017
145	<800	1578	1557	2244	2252
3758	3807	4607	4674	3503	3463
7693	7705	2619	2647	3871	3846
890	888	4876	4674	981	979
642	<800	1597	1557	296	<800
627	<800	2512	2476	82	<800
9294	9298	1003	975	4080	4017
937	948	720	741	1868	1846
1671	1662	3131	3124	2164	2150
390	<800	1591	1557	661	<800
1102	1107	1033	975	505	<800
467	<800	2457	2476	290	<800

7721	7705	4925	5025	5927	6248
1912	1880	5249	5271	7454	7495
57	<800	1965	1977	3187	3202
2809	2850	802	741	1416	1424
188	<800	2883	2897	1489	1424
4011	4052	523	<800	696	<800
48	<800	746	741	9782	9681
763	765	857	872	4027	4017
1336	1342	1915	1977	68	<800
1486	1454	1481	1439	5698	5840
1136	1107	6467	6607	2913	2925
5365	5335	671	<800	1999	2007
2191	2190	5011	5025	810	822
3418	3454	400	<800	50	<800
111	<800	3989	3787	2317	2252
1032	1040	7133	7253	4222	4154
3078	3067	1323	1309	3052	3036
4342	4359	4730	4674	665	<800
111	<800	1226	1213	801	822
224	<800	948	975	170	<800
8194	8205	275	<800	2028	2007
2340	2387	3763	3787	1306	1317
7439	7705	10576	10547	6252	6248
1471	1454	3799	3787	5603	5543
153	<800	13397	13418	2039	2007
5613	5590	98	<800	188	<800
1379	1342	4810	4674	766	<800
4087	4052	200	<800	582	<800
1261	1243	1998	1977	1340	1317
539	<800	926	975	683	<800
4416	4359	3824	3787	9003	8843
		5578	5601	6390	6248
		1834	1794	543	<800
		1374	1309	5270	5161
		1862	1794	832	822
		1949	1977	5561	5543
		878	975	186	<800

Query Match 2.2%; Score 21; DB 9; Length 166496;
 Best Local Similarity 100.0%; Pred. No. 5.3;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TATTGGAGAGGCATTAAATA 52

Db 24488 TATTGGAGAGGCATTAAATA 24468

RESULT 19

AC018846/c

LOCUS AC018846 175518 bp DNA linear PRI 29-MAR-2002

DEFINITION Homo sapiens chromosome 16 clone RP11-370P15, complete sequence.

AC018846

AC018846.4 GI:19807787

KEYWORDS

HTG.

SOURCE

ORGANISM

Homo sapiens.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 175518)

DOE Joint Genome Institute.

2 (bases 1 to 175518)

Sequencing of Human Chromosome 16

Unpublished

DOE Joint Genome Institute.

Direct Submission

Submitted (21-DEC-1999) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

3 (bases 1 to 175518)

DOE Joint Genome Institute.

Direct Submission

Submitted (05-MAR-2002) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

4 (bases 1 to 175518)

DOE Joint Genome Institute.

Direct Submission

Submitted (29-MAR-2002) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

FEATURES

source

1. 175518

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="16"

/clone="RP11-370P15"

BASE COUNT

57156 a 32486 c 31996 g 53880 t

Query Match

Best Local Similarity 100.0%; Pred. No. 5.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
misc_feature /note="assembly_name:Contig72"
26495..31534
/note="assembly_name:Contig73"
31635..37164
/note="assembly_name:Contig74"
37265..44921
/note="assembly_name:Contig75
clone_end:SP6
vector_side:right"
45022..50595
/note="assembly_name:Contig76"
50696..59103
/note="assembly_name:Contig77"
59204..69267
/note="assembly_name:Contig78"
69368..82673
/note="assembly_name:Contig79"
82774..97615
/note="assembly_name:Contig80"
97716..115178
/note="assembly_name:Contig81"
115279..128781
/note="assembly_name:Contig82"
128882..144470
/note="assembly_name:Contig83"
144571..161750
/note="assembly_name:Contig84"
161851..187038
/note="assembly_name:Contig85"
misc_feature /note="assembly_name:Contig85"
55977 a 36931 c 36406 g 55111 t 2613 others
BASE COUNT
ORIGIN
```

```
Query Match 2.2%; Score 21; DB 2; Length 187038;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 17 TTGAAGAGCAAAATATTGC 37
|||||
Db 37096 TTGAAGAGCAAAATATTGC 37076
```

```
RESULT 21
AC019228 196216 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 11, clone RP11-570N12, WORKING DRAFT
DEFINITION
AC019228 SEQUENCE, 20 unordered pieces.
ACCESSION AC019228.5 GI:8568940
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 196216)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 196216)
Waterston, R.H.
Direct Submission
Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 16, 2000 this sequence version replaced gi:7233983.
```

```
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0570N12
----- Summary Statistics -----
Sequencing vector: M13; 67%
```

Sequencing vector: plasmid; 33%
Chemistry: Dye-primer ET; 67% of reads
Chemistry: Dye-terminator Big Dye; 33% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 182091 bases at least Q40
Consensus quality: 186866 bases at least Q30
Consensus quality: 189092 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 194316; sum-of-contigs
Quality coverage: 3.88 in Q20 bases; sum-of-contigs
Quality coverage: 3.68 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 1722: contig of 1722 bp in length
* 1723 1832: gap of unknown length
* 1823 3457: contig of 1635 bp in length
* 3458 3557: gap of unknown length
* 3558 6192: contig of 2635 bp in length
* 6193 6292: gap of unknown length
* 6293 9327: contig of 3034 bp in length
* 9327 9427: gap of unknown length
* 9427 12746: contig of 3320 bp in length
* 12747 12846: gap of unknown length
* 12847 16489: contig of 3643 bp in length
* 16490 16589: gap of unknown length
* 16590 19714: contig of 3125 bp in length
* 19715 19814: gap of unknown length
* 19815 26224: contig of 6410 bp in length
* 26225 26324: gap of unknown length
* 26325 31243: contig of 4919 bp in length
* 31244 31343: gap of unknown length
* 31344 38835: contig of 7492 bp in length
* 38836 38935: gap of unknown length
* 38936 46701: contig of 7766 bp in length
* 46702 46801: gap of unknown length
* 46802 57002: contig of 10201 bp in length
* 57003 57102: gap of unknown length
* 57103 66636: contig of 9534 bp in length
* 66637 66736: gap of unknown length
* 66737 76277: contig of 9341 bp in length
* 76278 76377: gap of unknown length
* 76378 87128: contig of 10751 bp in length
* 87129 87228: gap of unknown length
* 87229 98549: contig of 11321 bp in length
* 98550 98650: gap of unknown length
* 98651 109492: contig of 10843 bp in length
* 109493 109592: gap of unknown length
* 109593 124679: contig of 15087 bp in length
* 124680 124779: gap of unknown length
* 124780 155821: contig of 31042 bp in length
* 155822 155921: gap of unknown length
* 155922 196216: contig of 40295 bp in length.
```

FEATURES

source

```
1..196216
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-570N12"
1..1722
/note="assembly_name:Contig8"
1823..3457
/note="assembly_name:Contig10"
3558..6192
/note="assembly_name:Contig11
clone_end:SP6
vector_side:right"
```

```

misc_feature      6293..9326
                   /note="assembly_name:Contig12"
misc_feature      9427..12746
                   /note="assembly_name:Contig13"
misc_feature      12847..16489
                   /note="assembly_name:Contig14"
misc_feature      16590..19714
                   /note="assembly_name:Contig15
clone_end::T7
vector_side:left"
misc_feature      19815..26224
                   /note="assembly_name:Contig16"
misc_feature      26325..31243
                   /note="assembly_name:Contig17"
misc_feature      31344..38835
                   /note="assembly_name:Contig18"
misc_feature      38936..46701
                   /note="assembly_name:Contig19"
misc_feature      46802..57002
                   /note="assembly_name:Contig20"
misc_feature      57103..66636
                   /note="assembly_name:Contig21"
misc_feature      66737..76277
                   /note="assembly_name:Contig22"
misc_feature      76378..87128
                   /note="assembly_name:Contig23"
misc_feature      87229..98549
                   /note="assembly_name:Contig24"
misc_feature      98650..109492
                   /note="assembly_name:Contig25"
misc_feature      109593..124679
                   /note="assembly_name:Contig26"
misc_feature      124780..155821
                   /note="assembly_name:Contig27"
misc_feature      155922..196216
                   /note="assembly_name:Contig28"
BASE COUNT      57961 a 35701 c 36544 g 64086 t 1924 others
ORIGIN
Query Match      2.2%; Score 21; DB 2; Length 196216;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 386 TATTTTGAATTCGATG 406
DB 63025 TATTTTGAATTCGATG 63045

RESULT 22
AC084217          216988 bp   DNA   linear   ROD 15-MAY-2002
LOCUS            Mus Musculus Strain C57BL/6J Chromosome 5 RP23-327P24, complete
DEFINITION      sequence.
ACCESSION      AC084217
VERSION        AC084217.7 GI:13549250
KEYWORDS
SOURCE
ORGANISM      Mus musculus.
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      1 (bases 1 to 216988)
               Montgomery,K.T., Grills,G., Han,J., Chiu,D., Decker,J., Fusina,M.,
               Goltz,J., Haider,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
               Perez,A., Shim,C., Thomas,E. and Kucherlapati,R.
TITLE        High Throughput Mouse Sequencing
JOURNAL
REFERENCE      2 (bases 1 to 216988)
AUTHORS      Montgomery,K.T., Grills,G., Han,J., Chiu,D., Decker,J., Fusina,M.,
               Goltz,J., Haider,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
               Perez,A., Shim,C., Thomas,E. and Kucherlapati,R.
TITLE        Direct Submission
JOURNAL
REFERENCE      Submitted (05-APR-2001) Department of Molecular Genetics, Albert
AUTHORS      Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
               Bronx, NY 10461, USA
               On Apr 5, 2001, this sequence version replaced gi:11177930.
COMMENT
               -----Genome Center:
               Harvard Partners Genome Center
               Center Code: HPGC
               Web site: http://www.hpgc.org/Sequence/mouse.html
               Contact: hpgc@mednet.mgh.harvard.edu

```

BRONX, NY 10461, USA
 3 (bases 1 to 216988)
 Montgomery,K.T., Grills,G., Han,J., Chiu,D., Decker,J., Fusina,M.,
 Goltz,J., Haider,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
 Perez,A., Shim,C., Thomas,E. and Kucherlapati,R.
 Direct Submission
 Submitted (05-APR-2001) Department of Molecular Genetics, Albert
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
 Bronx, NY 10461, USA
 On Apr 5, 2001, this sequence version replaced gi:11177930.
 -----Genome Center:
 Harvard Partners Genome Center
 Center Code: HPGC
 Web site: http://www.hpgc.org/Sequence/mouse.html
 Contact: hpgc@mednet.mgh.harvard.edu

CLONE LENGTH: This sequence represents the entire insert of this
 clone unless otherwise noted. If there are overlapping clones, the
 overlaps are noted in the beginning and end of the Features
 listing.

ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550).
 Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished.) for Human and Mouse sequences.
 Genes and Regions of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST
 and cDNA sequences in Unigene. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintain sequence continuity
 across the splice junctions. Sequences that are not identical
 matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double
 stranded sequence for all regions. All sequence is completed to a
 standard of coverage with a minimum of 3 reads with no ambiguities.
 If the sequence coverage for a region does not meet this standard,
 it is indicated in the annotation as Low Coverage. Low coverage
 linkages are verified by PCR product size verification or
 verification of forward and reverse reads from clones which span
 the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated average error rate is less than 1 per 10,000
 bases using the Consed quality parameters. Regions that do not
 meet this requirement are annotated as Low Quality.

-----Summary Statistics

Center project name: ACP
 Sequencing vector: pUC18; L08752
 Chemistry: Dye-terminator Big Dye; 100%
 Assembly program: Phrap version 0.990319
 Contig length: 216988
 Fraction of Phrap value < 40: 0.00216
 Error Rate in Consed: 0.02 per 10,000 bases
 Number of N's in consensus: 0

----- Distribution of Quality < 40 Bases:

#	bases	5	10	15	20	25	30	35	40
10001	7001	*	*	*	*	*	*	*	*
9001	6001	*	*	*	*	*	*	*	*
8001	5001	*	*	*	*	*	*	*	*
7001	4001	*	*	*	*	*	*	*	*
6001	3001	*	*	*	*	*	*	*	*
5001	2001	*	*	*	*	*	*	*	*
4001	1001	*	*	*	*	*	*	*	*
3001	0	*	*	*	*	*	*	*	*
2001									
1001									
0									

Phrap Value Range

FEATURES

URRS	Location/Qualifiers
source	1. 216988
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/chromosome="5"
repeat_region	/clone="RP23-327P24"
	complement(346..431)
repeat_region	/rpt_family="B1F"
	complement(440..531)
repeat_region	/rpt_family="LIMB4"
	1102..1250
repeat_region	/rpt_family="MIR3"
	complement(2900..3005)
repeat_region	/rpt_family="B3"
	complement(3011..3067)
repeat_region	/rpt_family="B2_Mm2"
	complement(3094..3186)
misc_feature	/rpt_family="RS1NE1"
	306..4211
	/note="Single Stranded"
misc_feature	4480..4586
	/standard_name="low coverage"
repeat_region	/note="Single Stranded"
	5079..5122
repeat_region	/rpt_family="CT_rich"
	5315..5351
repeat_region	/rpt_family="AT_rich"
	5687..5732
repeat_region	/rpt_family="(TTCA)n"
	complement(5747..5895)
repeat_region	/rpt_family="Lx7"
	5895..6467
repeat_region	/rpt_family="Lx6"
	6474..6562
repeat_region	/rpt_family="(GAA)n"
	6563..6928
repeat_region	/rpt_family="Lx7"
	complement(7082..7287)
repeat_region	/rpt_family="(URRIA"
	7848..8033
misc_feature	/rpt_family="B2_Mm2"
	7865..8040
	/standard_name="low coverage"
	/note="Single Stranded"
repeat_region	/rpt_family="(TA)n"
	8124..8178
repeat_region	8242..8456
repeat_region	/rpt_family="B3"
	complement(9238..9314)
repeat_region	/rpt_family="B1_MM"
	9306..9683
repeat_region	/rpt_family="MTA"
	9933..9996
repeat_region	/rpt_family="I03"
	11231..11286
repeat_region	/rpt_family="AT_rich"
	11460..11502
repeat_region	/rpt_family="(CTCA)n"
	complement(11556..11712)
repeat_region	/rpt_family="B2_Mm2"
	complement(11854..12056)
repeat_region	/rpt_family="B3A"
	12596..12666
repeat_region	/rpt_family="MTD"
	complement(13388..13581)
repeat_region	/rpt_family="B3"
	complement(13612..13691)
repeat_region	/rpt_family="IB3"
	13763..13989
repeat_region	/rpt_family="(URRIA"
	complement(15195..15584)
	/rpt_family="B3"

repeat_region	16403. .16409
/rpt_family="TA)n"	16410. .16782
repeat_region	/rpt_family="ORR1B"
repeat_region	16783. .16854
/rpt_family="TA)n"	16855. .16880
repeat_region	/rpt_family="(TCTA)n"
repeat_region	16901. .16962
/rpt_family="(CA)n"	complement(17348. .17504)
repeat_region	/rpt_family="B3"
repeat_region	18418. .18446
/rpt_family="(CAAA)n"	19111. .19130
repeat_region	/rpt_family="(TTTTTA)n"
repeat_region	19131. .19159
/rpt_family="(TG)n"	complement(19169. .19304)
repeat_region	/rpt_family="B3A"
repeat_region	complement(19374. .19477)
repeat_region	/rpt_family="PB1D7"
repeat_region	complement(20305. .20380)
/rpt_family="B1F"	/rpt_family="B1F"
repeat_region	complement(20520. .20642)
/rpt_family="B3"	complement(20643. .21039)
repeat_region	/rpt_family="MRB"
repeat_region	complement(21040. .21116)
/rpt_family="B3"	21484. .21507

Query Match	2.2%	Score 21	DB 10	Length 216988
Best Local Similarity	100.0%	Pred. No. 5.3		
Matches	21	Conservative	0	Mismatches 0
			Indels	0
			Gaps	0
QY	387	ATTTTGAATTCGATGA	407	
Db	99017	ATTTTGAATTCGATGA	99037	

RESULT	23			
LOCUS	AC125205	243381 bp	DNA	linear
DEFINITION	Mus musculus chromosome UTK clone RP23-23812, WORKING DRAFT			
ACCESSION	AC125205	SEQUENCE, 7 unordered pieces.		
VERSION	AC125205.1	GI:21490737		
KEYWORDS	HTG, HTGS_PHASE1, HTGS_DRAFT.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 243381)			
TITLE	McPherson,J.D. and Waterston,R.H.			
JOURNAL	The sequence of Mus musculus clone			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 243381)			
TITLE	McPherson,J.D. and Waterston,R.H.			
JOURNAL	Direct Submission			
COMMENT	Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA			

```

Center: Washington University Genome Sequencing Center
Center code: MUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BA0238102
Summary Statistics
Sequencing vector: MJ3; 0%

```


COMMENT OS (hepatitis C virus)
PN JP 1994000085-A/44
PD 11-JAN-1994
PR 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51, C12N15/70, C12N15/85, C12P21/02, C07K13/00, C07K15/12, PC
C12N1/21, C12N5/10,
C12N15/11, C12N15/70, C12N15/85, C12P21/02, A61K39/00, A61K39/29,
PC (C12N1/21),
PC C12R1.19), (C12N5/10, C12R1.91), (C12P21/02, C12R1.19), (C12P21/02,
PC C12R1.91),
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25-1;
FH Key Location/Qualifiers
FT mat_peptide 1..849
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient'.
FT Location/Qualifiers
FEATURES
source
1..849
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 146 a 244 c 249 g 210 t
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 CAGAAGACACAGGAAGA 330
Db 125 CAGAAGACACAGGAAGA 106
RESULT 26
E06205/c 849 bp RNA linear PAT 29-SEP-1997
LOCUS E06205
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06205.1 GI:2174392
VERSION JP 1994000085-A/45.
KEYWORDS Hepatitis C virus.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 849)
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
TITLE GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
JOURNAL CODED BY THE SAME AND ITS PRODUCTION
PATENT: JP 1994000085-A 45 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PD 11-JAN-1994
PD 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO

PC C12N15/51, C07K7/06, C07K7/08, C07K7/10, C07K13/00, C07K15/12, PC
C12N1/21, C12N5/10,
PC C12N15/11, C12N15/70, C12N15/85, C12P21/02, A61K39/00, A61K39/29,
PC (C12N1/21),
PC C12R1.19), (C12N5/10, C12R1.91), (C12P21/02, C12R1.19), (C12P21/02,
PC C12R1.91),
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25-2;
FH Key Location/Qualifiers
FT mat_peptide 1..849
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient'.
FT Location/Qualifiers
FEATURES
source
1..849
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 150 a 251 c 246 g 202 t
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 CAGAAGACACAGGAAGA 330
Db 125 CAGAAGACACAGGAAGA 106
RESULT 27
E06206/c 849 bp RNA linear PAT 29-SEP-1997
LOCUS E06206
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06206
VERSION E06206.1 GI:2174393
KEYWORDS JP 1994000085-A/46.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
VIRUSES: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 849)
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
TITLE GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
JOURNAL CODED BY THE SAME AND ITS PRODUCTION
PATENT: JP 1994000085-A 46 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PD 11-JAN-1994
PD 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51, C07K7/06, C07K7/08, C07K7/10, C07K13/00, C07K15/12, PC
C12N1/21, C12N5/10,
PC C12N15/11, C12N15/70, C12N15/85, C12P21/02, A61K39/00, A61K39/29,
PC (C12N1/21),
PC C12R1.19), (C12N5/10, C12R1.91), (C12P21/02, C12R1.19), (C12P21/02,
PC C12R1.91),
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;

CC *source: clone-MX25-3;
FH Key Location/Qualifiers
FT mat-peptide 1..849
FT /product='the peptides reacting specifically
FT and
FT immunohistochemically with the serum of hepatitis
FT type C
FT patient'
FT Location/Qualifiers
FEATURES
Source 1..849
/organism='Hepatitis C virus'
/db_xref='taxon:11103'
BASE COUNT 144 a 246 c 248 g 211 t
ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 311 CAGAGAACACAGGAAGA 330
Db 125 CAGAGAACACAGGAAGA 106

RESULT 28
E06389 849 bp RNA linear PAT 29-SEP-1997
LOCUS E06389
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06389
VERSION E06389.1 GI:2174576
KEYWORDS JP 1994000086-A/33.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C virus.
1 (bases 1 to 849)
Seki, M., Honda, Y., Takahashi, K., Murakami, T., Teranishi, Y. and
Hayashi, N.
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 33 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/33
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51, C07K7/06, C07K7/10, C07K13/00, C12N5/10, PC
C12N15/11,
PC C12N15/85, C12P21/02//A61K39/29, (C12P21/02, C12R1:91), C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25-1;
FH key Location/Qualifiers
FH mat-peptide 1..849
FT /product='the peptides reacting specifically
FT and
FT immunohistochemically with the serum of hepatitis
FT type C
FT patient'
FT Location/Qualifiers
FEATURES
Source 1..849
/organism='Hepatitis C virus'
/db_xref='taxon:11103'
BASE COUNT 132 a 227 c 232 g 188 t 70 others
ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 311 CAGAGAACACAGGAAGA 330
Db 125 CAGAGAACACAGGAAGA 106

RESULT 29
E06400 849 bp RNA linear PAT 29-SEP-1997
LOCUS E06400
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06400
VERSION E06400.1 GI:2174587
KEYWORDS JP 1994000086-A/44.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C virus.
1 (bases 1 to 849)
Seki, M., Honda, Y., Takahashi, K., Murakami, T., Teranishi, Y. and
Hayashi, N.
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 44 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/44
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51, C07K7/06, C07K7/10, C07K13/00, C12N5/10, PC
C12N15/11,
PC C12N15/85, C12P21/02//A61K39/29, (C12P21/02, C12R1:91), C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25-1;
FH key Location/Qualifiers
FH mat-peptide 1..849
FT /product='the peptides reacting specifically
FT and
FT immunohistochemically with the serum of hepatitis
FT type C
FT patient'
FT Location/Qualifiers
FEATURES
Source 1..849
/organism='Hepatitis C virus'
/db_xref='taxon:11103'
BASE COUNT 146 a 244 c 249 g 210 t
ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 311 CAGAGAACACAGGAAGA 330
Db 125 CAGAGAACACAGGAAGA 106

RESULT 30
E06401 849 bp RNA linear PAT 29-SEP-1997
LOCUS E06401
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06401

VERSION E06401.1 GI:2174588
KEYWORDS JP 1994000086-A/45.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 849)
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000086-A 45 11-JAN-1994;
COMMENT MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/45
PD 11-JAN-1994
PR 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC
C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25-2;
FH key Location/Qualifiers
FT mat_peptide 1..849
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient'.
FT Location/Qualifiers
FEATURES
source 1..849
BASE COUNT 150 a 251 c 246 g 202 t
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 311 CAGAGAACACAGGAAGA 330
IIIIIIIIIIIIIIIIIIII
Db 125 CAGAGAACACAGGAAGA 106
RESULT 31
LOCUS E06402 849 bp RNA linear PAT 29-SEP-1997
DEFINITION CDNA encoding genes derived from hepatitis C virus.
ACCESSION E06402
VERSION E06402.1 GI:2174589.
KEYWORDS JP 1994000086-A/46.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 849)
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000086-A 46 11-JAN-1994;
COMMENT MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/46

PD 11-JAN-1994
PR 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC
C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25-3;
FH key Location/Qualifiers
FT mat_peptide 1..849
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient'.
FT Location/Qualifiers
FEATURES
source 1..849
BASE COUNT 144 a 246 c 248 g 211 t
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 849;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 311 CAGAGAACACAGGAAGA 330
IIIIIIIIIIIIIIIIIIII
Db 125 CAGAGAACACAGGAAGA 106
RESULT 32
LOCUS E06200 1280 bp RNA linear PAT 29-SEP-1997
DEFINITION CDNA encoding genes derived from hepatitis C virus.
ACCESSION E06200
VERSION E06200.1 GI:2174387
KEYWORDS JP 1994000085-A/40.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE 1 (bases 1 to 1280)
AUTHORS Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.
TITLE GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION
JOURNAL Patent: JP 1994000085-A 40 11-JAN-1994;
COMMENT MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000085-A/40
PD 11-JAN-1994
PR 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C07K15/12, PC
C12N1/21,C12N5/10,
PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/29,
PC (C12N1/21,
PC C12R1:19),(C12N5/10,C12R1:91),(C12P21/02,C12R1:19),(C12P21/02,
PC C12R1:91),
PC C07K99:00;

[illegible]

FT	patient'	type C
FEATURES	Location/Qualifiers	
source	1..1280	
	/organism="Hepatitis C virus"	
	/db_xref="taxon:11103"	
BASE COUNT	222 a 383 c 379 g 296 t	
ORIGIN		
Query Match	2.1%; Score 20; DB 6; Length 1280;	
Best Local Similarity	100.0%; Pred. No. 24;	
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	311 CAGAGAACACAGAGAAGA 330	
Db	125 CAGAGAACACAGAGAAGA 106	
RESULT 34		
E06217/c	1280 bp RNA linear	PAT 29-SEP-1997
LOCUS		
DEFINITION	CDNA encoding genes derived from hepatitis C virus.	
ACCESSION	E06217	
VERSION	E06217.1 GI:2174404	
KEYWORDS	JP 1994000085-A/57.	
SOURCE	Hepatitis C virus.	
ORGANISM	Hepatitis C virus	
	viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.	
REFERENCE	1 (bases 1 to 1280)	
AUTHORS	Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and Hayashi,N.	
	GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS. POLYPEPTIDE CODED BY THE SAME AND ITS PRODUCTION	
JOURNAL	Patent: JP 1994000085-A 57 11-JAN-1994;	
	MITSUBISHI KASEI CORP	
COMMENT	OS (hepatitis C virus)	
	PN JP 1994000085-A/57	
	PD 11-JAN-1994	
	PF 11-JUN-1992 JP 1992194497	
	PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR	
	07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR	
	20-APR-1992 JP 92P 99957	
	PI SEKI MAKOTO, HONDA YOSHICAZU, TAKAHASHI KAZUNOBU, PI	
	MURAKAMI TOMOKO,	
	PI TERANISHI YUTAKA, HAYASHI NORIO	
	PC C12N15/51,C07K7/06,C07K7/08,C07K7/10,C07K13/00,C07K15/12, PC	
	C12N1/21,C12N5/10,	
	PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/00,A61K39/29,	
	PC (C12N1/21,	
	PC C12P1/19),(C12N5/10,C12P1/91),(C12P21/02,C12P1/19),(C12P21/02	
	PC C12P1/91),	
	PC C07K9/00;	
	CC strandedness: Double;	
	CC topology: Linear;	
	CC anti-sense: No;	
	*source: clone=MX25026B-1;	
	key Location/Qualifiers	
FT	mat_peptide	1..1278
FT		/product='the peptides reacting specifically
FT		and
FT		immunochemically with the serum of hepatitis
FT		type C
FT		patient'
FEATURES	Location/Qualifiers	
source	1..1280	
	/organism="Hepatitis C virus"	
	/db_xref="taxon:11103"	
BASE COUNT	219 a 378 c 387 g 296 t	
ORIGIN		
Query Match	2.1%; Score 20; DB 6; Length 1280;	

Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGACACAGAGAAGA 330
|||||
Db 125 CAGAGACACAGAGAAGA 106

RESULT 35
E06396/C 1280 bp RNA linear PAT 29-SEP-1997
LOCUS
DEFINITION cDNA encoding genes derived from hepatitis C virus.
E06396
ACCESSION
E06396.1 GI:2174583
VERSION
JP 1994000086-A/40.
KEYWORDS
Hepatitis C virus.
SOURCE
Hepatitis C virus.
ORGANISM
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE
1 (bases 1 to 1280)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
TITLE
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 40 11-JAN-1994;
JOURNAL
MITSUBISHI KASEI CORP
COMMENT
OS (hepatitis C virus)
PN JP 1994000086-A/40
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/08,C07K7/10,C07K13/00,C12N5/10, PC
C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25026;
FH Key Location/Qualifiers
FT mat_peptide 1..1278
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient'.
FT Location/Qualifiers
1..1280
FEATURES
source
1..1280
BASE COUNT 203 a 354 c 356 g 269 t 98 others
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 1280;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGACACAGAGAAGA 330
|||||
Db 125 CAGAGACACAGAGAAGA 106

RESULT 36
E06412/C 1280 bp RNA linear PAT 29-SEP-1997
LOCUS
DEFINITION cDNA encoding genes derived from hepatitis C virus.
E06412
ACCESSION
E06412.1 GI:2174599
VERSION
JP 1994000086-A/56.
KEYWORDS

SOURCE
Hepatitis C virus.
Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE
1 (bases 1 to 1280)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
TITLE
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 56 11-JAN-1994;
JOURNAL
MITSUBISHI KASEI CORP
COMMENT
OS (hepatitis C virus)
PN JP 1994000086-A/56
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734

RESULT 37
E06413/C 1280 bp RNA linear PAT 29-SEP-1997
LOCUS
DEFINITION cDNA encoding genes derived from hepatitis C virus.
E06413
ACCESSION
E06413.1 GI:2174600
VERSION
JP 1994000086-A/57.
KEYWORDS
Hepatitis C virus.
SOURCE
Hepatitis C virus.
ORGANISM
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE
1 (bases 1 to 1280)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
TITLE
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 57 11-JAN-1994;
JOURNAL
MITSUBISHI KASEI CORP
COMMENT
OS (hepatitis C virus)
PN JP 1994000086-A/57
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734

FEATURES
source
1..1280
BASE COUNT 222 a 383 c 379 g 296 t
ORIGIN
Query Match 2.1%; Score 20; DB 6; Length 1280;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGACACAGAGAAGA 330
|||||
Db 125 CAGAGACACAGAGAAGA 106

RESULT 37
E06413/C 1280 bp RNA linear PAT 29-SEP-1997
LOCUS
DEFINITION cDNA encoding genes derived from hepatitis C virus.
E06413
ACCESSION
E06413.1 GI:2174600
VERSION
JP 1994000086-A/57.
KEYWORDS
Hepatitis C virus.
SOURCE
Hepatitis C virus.
ORGANISM
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
REFERENCE
1 (bases 1 to 1280)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
TITLE
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 57 11-JAN-1994;
JOURNAL
MITSUBISHI KASEI CORP
COMMENT
OS (hepatitis C virus)
PN JP 1994000086-A/57
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734

PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51, C07K7/06, C07K7/10, C07K13/00, C12N5/10, PC
C12N15/11,
PC C12N15/85, C12P21/02//A61K39/29, (C12P21/02, C12R1:91), C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25026B-1;
FH key Location/Qualifiers
FT mat_peptide 1..1278
FT /product='the peptides reacting specifically
FT and
FT Immunohchemically with the serum of hepatitis
FT type C
FT patient',
FT
FT
FEATURES
source Location/Qualifiers
1..1280
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 219 a 378 c 387 g 296 t
ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 1280;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
|||||
DB 125 CAGAGAACACAGGAAGA 106

RESULT 38
E06203/c 3564 bp RNA linear PAT 29-SEP-1997
LOCUS E06203
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06203
VERSION E06203.1 GI:2174390
KEYWORDS JP 1994000085-A/43.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
vitruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 3564)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000085-A 43 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000085-A/43
PD 11-JAN-1994
PF 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51, C07K7/06, C07K7/10, C07K13/00, C07K15/12, PC
C12N1/21, C12N5/10,
PC C12N15/11, C12N15/70, C12N15/85, C12P21/02//A61K39/00, A61K39/29,
PC (C12N1/21,
PC C12R1:19), (C12N5/10, C12R1:91), (C12P21/02, C12R1:19), (C12P21/02,
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;

CC anti-sense: No;
CC *source: clone-MX25N15;
FH key Location/Qualifiers
FH mat_peptide 1..3564
FT /product='the peptides reacting specifically
FT and
FT Immunohchemically with the serum of hepatitis
FT type C
FT patient',
FT
FT
FEATURES
source Location/Qualifiers
1..3564
/organism="Hepatitis C virus"
/db_xref="taxon:11103"
BASE COUNT 640 a 1019 c 1024 g 747 t 124 others
ORIGIN

QY 311 CAGAGAACACAGGAAGA 330
|||||
DB 125 CAGAGAACACAGGAAGA 106

RESULT 39
E06223/c 3564 bp RNA linear PAT 29-SEP-1997
LOCUS E06223
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06223
VERSION E06223.1 GI:2174410
KEYWORDS JP 1994000085-A/63.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
vitruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 3564)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000085-A 63 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000085-A/63
PD 11-JAN-1994
PF 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51, C07K7/06, C07K7/10, C07K13/00, C07K15/12, PC
C12N1/21, C12N5/10,
PC C12N15/11, C12N15/70, C12N15/85, C12P21/02//A61K39/00, A61K39/29,
PC (C12N1/21,
PC C12R1:19), (C12N5/10, C12R1:91), (C12P21/02, C12R1:19), (C12P21/02,
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25N15-1;
FH key Location/Qualifiers
1..3564
FT /product='the peptides reacting specifically
FT and
FT Immunohchemically with the serum of hepatitis
FT type C
FT patient',
FT

FEATURES
source 1. .3564
Location/Qualifiers
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT 664 a 1062 c 1055 g 783 t

ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 3564;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
|||||

Db 125 CAGAGAACACAGGAAGA 106

RESULT 40
E06399/c
LOCUS E06399 3564 bp RNA linear PAT 29-SEP-1997
DEFINITION CDNA encoding genes derived from hepatitis C virus.
ACCESSION E06399.1 GI:2174586
VERSION JP 1994000086-A/43.
KEYWORDS JP 1994000086-A/43.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
Hepatitis C virus.
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 3564)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 43 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/43
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC
C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25N15;
FH Key Location/Qualifiers
FH mat_peptide 1. .3564
FT /product="the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient".
FT Location/Qualifiers
1. .3564
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT 640 a 1019 c 1024 g 747 t 134 others

ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 3564;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
|||||

Db 125 CAGAGAACACAGGAAGA 106

RESULT 41
E06419/c
LOCUS E06419 3564 bp RNA linear PAT 29-SEP-1997
DEFINITION CDNA encoding genes derived from hepatitis C virus.
ACCESSION E06419.1 GI:2174606
VERSION JP 1994000086-A/63.
KEYWORDS JP 1994000086-A/63.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
Hepatitis C virus.
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 3564)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 63 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/63
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/10,C07K13/00,C12N5/10, PC
C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone-MX25N15-1;
FH Key Location/Qualifiers
FH mat_peptide 1. .3564
FT /product="the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient".
FT Location/Qualifiers
1. .3564
/organism="Hepatitis C virus"
/db_xref="taxon:11103"

BASE COUNT 664 a 1062 c 1055 g 783 t

ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 3564;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
|||||

Db 125 CAGAGAACACAGGAAGA 106

RESULT 42
AF313916/c
LOCUS AF313916 9359 bp mRNA linear VRL 01-JAN-2002
DEFINITION Hepatitis C virus polyprotein mRNA, partial cds.
ACCESSION AF313916
VERSION AF313916.1 GI:18027684
KEYWORDS
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
Hepatitis C virus.
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
1 (bases 1 to 9359)
Fanning,L.J., Itakura,J., Nagayama,K. and Enomoto,N.
Characteristics of Hepatitis C viral genome associated with disease

PVPSDMETKITTWGADTAACGDIILGLPVSARCKEILLGPADSLGQGMRLAPIT
 AYSQTRKLGCIITSLGRKNQVEGEVAVSTQSFALCINGVCTVYHGKSK
 TLACGKCPITQWTVNDDVLGMPAPGASMTPTCCSGTILVTRADVYVARRG
 DSRGSLSPRISTYLGKSGGGLPCSGHVVAIFRAACTRGVAVPPEMSEET
 MRSPTFTDNTSTPPAYQSTFOVAHLAHPGSGSKSTVPAAVAGKIVLVNVAATL
 GGAATMSKAHGVDPNIRTVGRTITTGAPITVSTYKFLADGCGSGADVLIICDCCHS
 TDSTSLIGTVLDOAETAGARLVLAATAPGSAVTPHPNIEVALSTGTEIPYGA
 AIPLEAIGGRHLIFCHSKKCDLEAKLSAGVANAVAYRLDVSIIPTSGDVVVA
 TDLMTGYTGDSDVSDICNTCVOTFVDSLDPTFTEITTYTODASVORGRFGRG
 RGGIATRYTPGRSGMFDSSVLCCECYAGACMVELTAENSVRIRATVINTPQVCO
 DHEFWESYFTGLTHIDAHFLSQTQACDNFYLVAQATVCARAKAPPSSDMQMKC
 LIRLKPPLHGPPLLYRLGAVQNEVTLHPITKFLMACMSADELVSTWVLVGVLA
 ALAACTLTGTVIVIGRIILSGRPVVIDPREVLXOEPMEECASHLPYIEGQMLAE
 OFKOKALILQATKOAEAAAVVSESKRALETTPAKMMHNTISGTOYLAGISTLPN
 PAIASIMAFATISITPLTTOYTLLENIIGWVAADLAPPASAFVAGAGAAVGS
 GUGKYVDIILAGYGAGVAGALVAFKVMGSDMSTEDVNLPAIISPCALVYGVCA
 ILRRVYGEBAVQMMNRLIFASRGNHVSPHYVESDAARVTOIISNLTITOLK
 RLHOWINDECSPTSCGSMRLDWMICVYLADFKTWSLKLPLRGVFFESCQYK
 GVRGDIIMYITPCGAQITGHVKNKSMRIVGPRCSNTMGTPTINATYTGCTPSP
 APNYSRALMRVAAEEVETRYGDFHYVTGMTDNKPCQVPAPEFTELDGVILHR
 YAPACKPLRDEVTFOVGLNOYTVGSQLCEPEPVTYVTSMLTDPHSHTAAARRL
 ARGSPSLASSASAOLSLKATCTTHGAPDPLITANILMROMGNTIRVSEN
 KIVIIDSEPLAEEDEREVVAATILKTRKFPAMVWABPDINPPLLESKAPDY
 VPPVHGGCPLEPTKAPPIPPRRKRTVVLTSTVSSALAEATKTFGSSGSAVDST
 ATGPDQASABGDAGSDAESYSMPLLEGEPDPLNDGSMSTVSEASSEDVCCMS
 YTMGTALITPCAAEESKLPINALSNPLLRHNMVATYRSASOROKKYTFPRLVD
 DHPDVEYDELEITSCSSNVVAHDASGRVYLLDPTPTPLARAMEETARHPVNSML
 GNIIMVAPTLMARMLTMHFFSIIILAOOLEKALDIOYGAATYSEIPLDPIIORLH
 GLSAFSLSYSGEINRYASOCLRKIGVPLRYWRBARSVRAKILISOGGRATCKYL
 FNMAYKTKLKLPIPEASQDLSGMFVAGYSGGDIYHLSLRARPPRMFMCLLLSVY
 GYILPNR"

misc_feature
 3029..4490
 /note="Csp1-XhoI(MKc1a clone), other sequence(MKc1c
 clone)"

BASE COUNT 1893 a 2820 c 2693 g 2025 t

ORIGIN

Query Match 2.1%; Score 20; DB 14; Length 9431;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACAACAAGAAAGA 330
 |||||||||||||||||||
 Db 2659 CAGAGAACAACAAGAAAGA 2640

RESULT 44
 HPCP/c 9431 bp RNA linear VRL 07-FEB-1999

LOCUS Hepatitis C virus complete genome sequence.
 DEFINITION D30613
 VERSION D30613.1 GI:560788
 KEYWORDS polyprotein.
 SOURCE Hepatitis C virus cDNA to genomic RNA.
 ORGANISM Hepatitis C virus
 Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 1 (sites)
 AUTHORS Seki,M., Honda,Y., Kondo,J., Fukuda,K., Ohta,K., Sugimoto,J. and
 Yamada,E.
 TITLE Effective production of the hepatitis C virus core antigen having
 high purity in *Escherichia coli*
 JOURNAL J. Biotechnol. 38 (3), 229-241 (1995)
 MEDLINE 95186115
 REFERENCE 2 (bases 1 to 9431)
 AUTHORS Seki,M. and Honda,Y.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 9431)

AUTHORS Seki,M.
 TITLE Direct Submission
 JOURNAL Submitted (29-Apr-1994) Makoto Seki, Mitsubishi Chemical
 Corporation Yokohama Research Center, Department of Biotechnology,
 Laboratory, 1000 Kamoshida-cho, Aoba-ku, Yokohama, Kanagawa 227,
 Japan (E-mail:seki@atlas.ic.m.kagaku.co.jp, Tel:045-963-3455,
 Fax:045-963-3992)
 COMMENT The serums of a chronic HCV patients were gifted by Dr. N. Hayashi
 and Dr. E. Mita(Osaka University Medical School).
 FEATURES
 source Location/Qualifiers
 1..9431
 /organism="Hepatitis C virus"
 /db_xref="taxon:11103"
 1..341
 /db_xref="GI:560789"
 342..9374
 /codon_start=1
 /product="polyprotein"
 /protein_id="BAA06303.1"
 /translation="MSTNPKPQKIKRNTNRRODVKKPPGGQIVGVYLLPRGRPL
 GYVATRKTSERPQGRGROPKAPQREGRAAGQYWPVLTGNGELGAGGLLSPG
 SNRSGPTDPRRSNRLKVIDTLGCGFDLMGTPYAGPILGGAARALABRVYLED
 GNVATGNLPGCSFIFLALISCLTIPASAVQVNRASGVYHTDCCSSSTVYRAD
 VYMHDPGCVNRENNSRCVWALPTLAARNSLPTTILRRHVDLVGAALCSAMY
 VGDPCGYELVSQLETFPSRREYVQDCSTYPGVSGHRAMAMIMMNSPTALVY
 SOLLRIPQAVDMVAGAHMGLAGLAAYSQVNMMAKVLVLMILFAGVDGCTVTGKV
 ATTQGFSTFESRSPSOKTQVNTNGSMHINTALNCDSLNTGFLAALFYHSPNAS
 CGPRMACRGITDEFAQCGMTITVYVNPISDQRPYCMHAPRPGIVPASYQCYVC
 ETPSPVYVGTDRFGAPTYNMGNTEDVLLNNTNRPQGMNKGCTMMNSTGFKTGG
 PCPNIGVGNNTLPDTCERKHPBATYTKCGSGMPLRCLVHPYRLMHPKCVNF
 TIFKVMYGVGHEHLEACNWTREKRDLEDRDSELSPLISTEOMVLCSEFTLL
 PALSTGLIHLHONIIVDQYLYIGSAVSFAIKWEYILLPLFLDARCAMLMMLL
 IAOAAALAEVNLVNAASAGAGVGLVLFVFCFAMVYIKGRVILPGAAVAYVYVMEH
 LLMAALPARYANDREMAASGCAVGVGLVLTSPYKFKLKLMLVQYLTTPRAH
 LOVWIPPLNRCGRATILITCAVHPELFDITKLALLGLGLVLMVQGLTQMPYTR
 AOGILRMCLVRAAGHYVOMALKLAALTSTYVYDHLTPLDMAHALRDLAAVE
 PVFSDMETKITTWGAETAACGDIISLSPVSARREGLILGPADEFGQGMRLAPIT
 AYSQTRKLGCIITSLGRKNQVEGEVAVSTQSFALCINGVCTVYHGKSK
 TLACGKCPITQWTVNDDVLGMSAPGASRLTPTCCSGSDIYVTRADVYVARRG
 DSRGSLSPRISTYLGKSGGGLPCSGHVVAIFRAACTRGVAVPPEMSEET
 MRSPTFTDNTSTPPAYQSTFOVAHLAHPGSGSKSTVPAAVAGKIVLVNVAATL
 GGAATMSKAHGVDPNIRTVGRTITTGAPITVSTYKFLADGCGSGADVLIICDCCHS
 TDSTSLIGTVLDOAETAGARLVLAATAPGSAVTPHPNIEVALSTGTEIPYGA
 AIPLEAIGGRHLIFCHSKKCDLEAKLSAGVANAVAYRLDVSIIPTSGDVVVA
 TDLMTGYTGDSDVSDICNTCVOTFVDSLDPTFTEITTYTODASVORGRFGRG
 RGGIATRYTPGRSGMFDSSVLCCECYAGACMVELTAENSVRIRATVINTPQVCO
 DHEFWESYFTGLTHIDAHFLSQTQACDNFYLVAQATVCARAKAPPSSDMQMKC
 LIRLKPPLHGPPLLYRLGAVQNEVTLHPITKFLMACMSADELVSTWVLVGVLA
 ALAACTLTGTVIVIGRIILSGRPVVIDPREVLXOEPMEECASHLPYIEGQMLAE
 OFKOKALILQATKOAEAAAVVSESKRALETTPAKMMHNTISGTOYLAGISTLPN
 PAIASIMAFATISITPLTTOYTLLENIIGWVAADLAPPASAFVAGAGAAVGS
 GUGKYVDIILAGYGAGVAGALVAFKVMGSDMSTEDVNLPAIISPCALVYGVCA
 ILRRVYGEBAVQMMNRLIFASRGNHVSPHYVESDAARVTOIISNLTITOLK
 RLHOWINDECSPTSCGSMRLDWMICVYLADFKTWSLKLPLRGVFFESCQYK
 GVRGDIIMYITPCGAQITGHVKNKSMRIVGPRCSNTMGTPTINATYTGCTPSP
 APNYSRALMRVAAEEVETRYGDFHYVTGMTDNKPCQVPAPEFTELDGVILHR
 YAPACKPLRDEVTFOVGLNOYTVGSQLCEPEPVTYVTSMLTDPHSHTAAARRL
 ARGSPSLASSASAOLSLKATCTTHGAPDPLITANILMROMGNTIRVSEN
 KIVIIDSEPLAEEDEREVVAATILKTRKFPAMVWABPDINPPLLESKAPDY
 VPPVHGGCPLEPTKAPPIPPRRKRTVVLTSTVSSALAEATKTFGSSGSAVDST
 ATGPDQASABGDAGSDAESYSMPLLEGEPDPLNDGSMSTVSEASSEDVCCMS
 YTMGTALITPCAAEESKLPINALSNPLLRHNMVATYRSASOROKKYTFPRLVD
 DHPDVEYDELEITSCSSNVVAHDASGRVYLLDPTPTPLARAMEETARHPVNSML
 GNIIMVAPTLMARMLTMHFFSIIILAOOLEKALDIOYGAATYSEIPLDPIIORLH
 GLSAFSLSYSGEINRYASOCLRKIGVPLRYWRBARSVRAKILISOGGRATCKYL
 FNMAYKTKLKLPIPEASQDLSGMFVAGYSGGDIYHLSLRARPPRMFMCLLLSVY
 GYILPNR"

BASE COUNT

1882 a 2819 c 2701 g 2029 t

ORIGIN

Query Match 2.1%; Score 20; DB 14; Length 9431;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
|||||
DB 2659 CAGAGAACACAGGAAGA 2640

RESULT 45

E06261/c

LOCUS E06261 9471 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06261.1 GI:2174448
VERSION JP 1994000085-A/101.
KEYWORDS Hepatitis C virus.
SOURCE Hepatitis C virus.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

REFERENCE

1 (bases 1 to 9471)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
GENE OR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000085-A 101 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000085-A/101
PD 11-JAN-1994
PF 11-JUN-1992 JP 1992194497
PR 11-JUN-1991 JP 91P 139268, 12-JUL-1991 JP 91P 172794, PR
07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/08,C07K7/10,C07K13/00,C07K15/12, PC
C12N1/21,C12N5/10,
PC C12N15/11,C12N15/70,C12N15/85,C12P21/02//A61K39/00,A61K39/29,
PC (C12N1/21,
PC C12R1:19),(C12N5/10,C12R1:91),(C12P21/02,C12R1:19),(C12P21/02,
PC C12R1:91),
PC C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone=T7N1-30;
FH Key Location/Qualifiers
FH mat-peptide 369..9398
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient',
FT

COMMENT

JOURNAL

FEATURES

Source

BASE COUNT 1892 a 2830 c 2712 g 2037 t
ORIGIN

Query Match 2.1%; Score 20; DB 6; Length 9471;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGAACACAGGAAGA 330
|||||
DB 2686 CAGAGAACACAGGAAGA 2667

RESULT 46

E06457/c

LOCUS E06457 9471 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding genes derived from hepatitis C virus.
ACCESSION E06457
VERSION E06457.1 GI:2174644
KEYWORDS JP 1994000086-A/101.
SOURCE Hepatitis C virus.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

REFERENCE

1 (bases 1 to 9471)
Seki,M., Honda,Y., Takahashi,K., Murakami,T., Teranishi,Y. and
Hayashi,N.
GENE FOR DNA FRAGMENT DERIVED FROM HEPATITIS C VIRUS, POLYPEPTIDE
CODED BY THE SAME AND ITS PRODUCTION
Patent: JP 1994000086-A 101 11-JAN-1994;
MITSUBISHI KASEI CORP
OS (hepatitis C virus)
PN JP 1994000086-A/101
PD 11-JAN-1994
PF 07-OCT-1992 JP 1992293734
PR 07-OCT-1991 JP 91P 287008, 16-DEC-1991 JP 91P 332329, PR
20-APR-1992 JP 92P 99957
PI SEKI MAKOTO, HONDA YOSHIKAZU, TAKAHASHI KAZUNOBU, PI
MURAKAMI TOMOKO,
PI TERANISHI YUTAKA, HAYASHI NORIO
PC C12N15/51,C07K7/06,C07K7/08,C07K7/10,C07K13/00,C12N5/10, PC
C12N15/11,
PC C12N15/85,C12P21/02//A61K39/29,(C12P21/02,C12R1:91),C07K99:00;
CC strandedness: Double;
CC topology: Linear;
CC anti-sense: No;
CC *source: clone=T7N1-30;
FH Key Location/Qualifiers
FH mat-peptide 369..9398
FT /product='the peptides reacting specifically
FT and
FT immunochemically with the serum of hepatitis
FT type C
FT patient',
FT

COMMENT

FEATURES

source

BASE COUNT 1892 a 2830 c 2712 g 2037 t
ORIGIN

QY 311 CAGAGAACACAGGAAGA 330
|||||
DB 2686 CAGAGAACACAGGAAGA 2667

RESULT 47

AB049099/c

LOCUS AB049099 9504 bp RNA linear VRL 22-AUG-2002
DEFINITION Hepatitis C virus gene for polyprotein, complete cds,
Isolate:HCVT212.
AB049099
ACCESSION AB049099.1 GI:11559464
VERSION
KEYWORDS Hepatitis C virus (isolate:HCVT212, specific_host:human) serum cDNA
to genomic RNA.
SOURCE Hepatitis C virus
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

REFERENCE

1
Hepacivirus.

AUTHORS
Takahashi, K., Iwata, K., Matsumoto, M., Matsumoto, H., Nakao, K.,
Hatahara, T., Ohta, Y., Kanai, K., Maruo, H., Baba, K., Hijikata, M. and
Mishiro, S.

TITLE
Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
with hepatocellular carcinoma: the 'progression score' revisited

JOURNAL
Hepatol. Res. 20 (2), 161-171 (2001)

PMID
11348851

REFERENCE
2 (bases 1 to 9504)

AUTHORS
Mishiro, S.

TITLE
Direct Submission

JOURNAL
Submitted (19-SEP-2000) Shunji Mishiro, Toshiba General Hospital,
Department of Medical Sciences; 6-3-22 Higashi Oh-1, Shingaga-ku,
Tokyo 140-8522, Japan (E-mail: shunji.mishiro@po.toshiba.co.jp,
Tel: 81-3-3764-8981, Fax: 81-3-3764-8992)

FEATURES
Location/Qualifiers
1. 9504

source
/organism="Hepatitis C virus"
/isolate="HCV1212"
/specific_host="human"
/db_xref="taxon:11103"
/tissue_type="serum"
/note="HCV-genotype: 1b"
1. 341
/note="complete"
342. 9377
/codon_start=1
/product="polyprotein"
/protein_id="BAB1812.1"
/db_xref="GI:11559465"
/translation="MSTNPKPQRTKRNTRNRDQVCPKPGGQIVGVYLLPRGPRP
GVRATKTSERQPRGRORIPKARRPGRRAAGQGYWPLVGNMGMAWAGLLEPRG
SNRSGPTDPRRRRNLRKVIDLTGCGADLGLAIPYGAARALAHGVRVLE
GVNATGMLPGCSFISFLALISCLTIPASAEVNSGGYVNTDNCSSSYETAD
MIMHPCVPCVRESNSKRCWALPPTLARNNSVPTMIRHIDLVGTARFCSAM
VGDGGSYVNSQLTFPSRRHETVDDNCSTYSPHILGHRAAMMMAMSTALVY
SOLKIPQAVDMVAGMAGVLAGLAVYSMGNAKVALVLLFAGVGRIGLIGTT
SLNTRVASIESPSPQKIQLVNINGSWMINETALDCKASLQTLGAGLVAKRNTS
GCLERMASRPIDTPAOCMGPTTYTERVNSDQPYCMHAPRCGIIIPASQVCGVYC
FTSPVVVGTDLHGAPTSVRNGENTVLLNTPRPOGMMXCTNMNSTGTTKCGG
PPCNIGGGGNNLTCPDCEPKRHEATYTCGSGQIPRLVNYPRILMYPCPTN
FTIEKVMYVGVGHEHRSVACMWTBGRNCLEDRRSELSPLSTTQMILPCGFTT
LPALSTGLIHQGGIVDVOYLIGISAFVSFATKEVYLLLELLADRVCAQWML
LIAQEALENLVNLAASVAGAHILSFVFCAMYIKGLVGAAYALVAPL
LLALLPRAVAMREMAASCGAVFGLAEFLTSPYKVFARLIMLOVETRAEA
HLVWVPLNVGGRDAILLTCAIHPLEIDITLLAILGLPLVWLAGTAPEYFV
RAQGLIRACMLRVKAVGGHYOMALYKLAALITGYVYDHPIDOMAHVALBDLAVY
EPVPSDMETKLTITGADTPACGDIILGLPSARKKELLGRPASLEGCRLLAPI
TAYSQTRGLACITTSITGRDNQVBEVGVSTATQSGSLATYNGVCTVYHAGS
KTLAPKGPITQMTNVDQDLGWLAPPGARSLPTCTGSSDLVYTRHADIVPARR
GDGRSLSPRVSYLKGSSGGPLCPGSHVVGIRAVACITGVAKADVPEVESMT
TMRSPVFTDNSPSPAVPOTFOYVHLHAPTSGSKRVAAVYAGGVKVLVNPVSAAT
LSFGAYMSKAGTGNIRTVGVTITTGAPITGYSTYGRKLADGCGSGAYDITCECH
STDSPTIIGIGVLDQAEAGARLVVLAATPPGVTYVPHPIEVALSNTGEVFFG
KAIPLEATKGGHLLFCHSKKCYDELAKLSALGNAAVYKGLDVSVIPNGDVVY
ADALMTGFTGDSVIDCNCVTQVDFSLDPTPTITTTVPDVAVSQGRGTGR
GRRGYREVTBEPBSGMDSVLCEYDAGCAWELPAETSIIRAVLWNPGLPAC
ODHLEFMEGVFTGLTHIDAHFLSOTKOAGDNPEYVAVOATYCAARQAPSPSMDMMK
CLIRLKPHTGPTLLYRLGAVONELIILHPTTKTMCMSADLEVMVSTVYVGLV
AALAYCITGGSVIVGRILISGRVAVPDEVLYLREDEMECHSHLPYIDQKQLA
EFOKQALGLQDITAKQAEAAVAVVESWQALPWAHMMWFISGIDYLAGSLTPG
NPALASMAFASITSPLTQSTLLFNILGVAALAPASWAGVAGVAGVAGV
IGLRVLDIILAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGV
ALRRHVGGEGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGV
KLRHOMINEDSTPCSGSLRDVMDMCTVLTDEMTYQSLKPLRLPVPFPSCORGY
KGNRGEINOTITPCGNOIGHYKNGSMRIVPCTGCSNMVHGPIPIAATYTGPTPS
PAPNTSKALMVAEEVETRVGDHFVHTVGTQDITDNCVQVAPPEFTTDEGVRLS
RYAPACKLIDDEVSVFQVGLQVYLSQLEPEBPAVAVLSMLMDQPMIAETARR
LARGSPPLASSASSOLSPASLAKATCTHDSPADLIEANLMDQEGMNTRESE
NKVYVLDISDFPLRAEDEREVSVAPEILKRSKGPAPAIPIANRPYNPPLSEWDC
YVPPVYHGGCLPLTRAPPITPPRRKRTVLTSTESVSALEATCTSSSSSSANDP
TATAPPDQASDGDKESVSESSAPPLEGEPDGLDSDGSKSTVSEASEDEVCSSM
STYMGALITVPCAEESKLPINALNSLILRNHNHVAATYSASILRQKRVFDRLOVL
DDHRDVLKEMKAKASTVAKRLISVEACKLTPPHSAKSKFGYAKADNLSRAVNH

IRSVKDLLEDITVPIDITTVMAKNEVFCVQPEKGRKPARLIVPFDLGVCEKMAKY
DVASTLPQAVMGSSYGFQYSPGQVPEFLVFNAMKSKCPGDFGDFSTVTENDIR
VEESIXOCDDLSPEAROAKSLRTERLYIGPLTNSKGNCGVRCRAGSVLITSCGNT
LTCYKASACRAAKLODCMVLNDDLVYICDSNGVQDEACLAAPFAMRYRISAP
GDPPOPAIDLEITSCSSNVSAHDSGRVYTLTRDPTPLARAMEFARTTPVNSP
LGNITMVAETLWVRLMIMTHFESILLAEQLEKALDQIYGACYSIEPLDPOILERL
HLSAFSLHSYSPGEINRVASCLARKGVPELIMHRRARSVRKALSDGGRATCGKY
LENMAVKTIKLKLPIPAASOLDLSCWVAGVSGDITVHSLSRARRMFWLCLLSVG
VGVYLLPNR"

9378. 9415
/note="upstream of poly-u stretch"
9416. 9504
/note="poly(U) length is indistinct
poly(U) stretch"

BASE COUNT
1886 a 2810 c 2683 g 2125 t

ORIGIN

Query Match 2.1%; Score 20; DB 14; Length 9504;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 CAGAGACACACAGGAAGA 330
|||||
Db 2662 CAGAGACACACAGGAAGA 2643

RESULT 48
AE008634/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
1 (sites)
Ogata, H., Audic, S., Barbe, V., Artiguenave, F., Fournier, P. E.,
Raoult, D. and Claverie, J. M.
Selfish DNA in protein-coding genes of *Rickettsia*
Science 290 (5490), 347-350 (2000)

JOURNAL
MEDLINE
20485642

PUBMED
11030655

REFERENCE
2 (sites)
Ogata, H., Audic, S. and Claverie, J. M.
Selfish DNA and the origin of genes
Science 291 (5502), 252-253 (2001)

JOURNAL
MEDLINE
3 (bases 1 to 10435)

AUTHORS
Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P. E., Barbe, V.,
Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J. M. and
Raoult, D.

TITLE
Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*

JOURNAL
Science 293 (5537), 2093-2098 (2001)

PUBMED
11557893

REFERENCE
4 (bases 1 to 10435)
Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P. E., Barbe, V.,
Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J. M. and
Raoult, D.

TITLE
Direct Submission

JOURNAL
Submitted (26-APR-2001) Unite des Rickettsies, Faculte de Medecine,
CNRS UMR 6020, Universite de la Mediterranee, 27 Boulevard Jean
Moulin, Marseille Cedex 05 13385, France
A public version of *R. conorii* genome database is accessible at
<http://igs-server.cns-mrs.fr/>. The database intends to provide
updated data. Annotation of the genome is an ongoing task whose
goal is to make the genome sequence more useful. Comments to the
authors are appreciated.

FEATURES
source
1. 10435
/organism="Rickettsia conorii"

g115458267|emb|CAB49756.1|' (A248285) methylalonyl-COA
mutase, subunit alpha, N-terminus (mcmA1) [Pyrococcus
abyssi], score 762, E-value 0"
/codon_start=1
/transl_table=11
/product="Methylalonyl-COA mutase, N-terminal
domain/subunit
/protein_id="AAM25528.1"
/db_xref="GI:20517397"
/translation="MYDEKLEETIMKEDMEKNHKGAKAFPERKEEFTSEIEI
KPIYEDVVDKIDYLEKGPPEYPTGVOPTWVKGFTMRQVAGYGAESNKR
KYLEEGGTGVSVAFDLPOTGSDHMGAGKGVADISLYDMLFDGIDPK
VTSMTIAPAILAMTIAVAEKQNSPDNLKNTQNDITIEYVARTGTFPTGSM
RLTINIEYCSKVPKMTISISGTHMEGATVQVAFLLANAVDAIAKGLD
VDFAPRISPFSSANDLFEVFAFRARVMWAKIMKRFCAKPRSMRHHQTAG
STLAQCPDNRIIVTIOALAAVLGTSGLSHNSKDELAPTEDSVIALTQOIIA
VESGVDTPDLASYYEYLTDEKRAFYIKIDELGAVAIADGYMORIOES
AYRQMETESGRITVGNKPMEEPPKNIKVDPOVEIYQREKIKVKEMDNARY
EKALKAKKAEGDDNMPILDAVREATIGETIDVLSVFGEEQOQIIF"
complement(1861..3414)
/gene="Sbm5"
/note="Pfam match to entry M4_COA_mutase,
Methylalonyl-COA mutase, score 938.4, E-value 1.90E-278"
complement(3531..3535)
/gene="Sbm5"
complement(3608..3625)
/note="putative rho-independent transcription terminator"
complement(3625..4013)
/gene="TTE2390"
complement(3625..3990)
/gene="TTE2390"
/note="Best Blastp hit = g117434028|p|H64673
dihydropterolate synthase (EC 2.5.1.15) [similarity] -
Helicobacter pylori (strain 26695)
g112314394|gb|AAD08276.1| (AE000628) dihydropterolate
synthase (f01P) [Helicobacter pylori 26695], score 52,
E-value 2.00E-06"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM25529.1"
/db_xref="GI:20517398"
/translation="MRREVLKIDNLSEAKRELSQLKEEISVDIMAPKAVRVKLYD
VHPAANIYKQMLAAGAAVARGCVNNSVSDVIIMGTLKQYRRLSLKMKQSY
FGISIVKELEVEREMSK"
complement(4000..5679)
/gene="MIS1"
/note="TTE2391"
complement(4000..5667)
/gene="MIS1"
/note="Best Blastp hit = g1120563|sp|P21164|FTHS_MOOTH
FORMATE-TETRAHYDROFOLATE LIGASE (FORMTLETERAHYDROFOLATE
SYNTHETASE) (FHS) (FTHFS) g1198640|p|I1A35942
formate-tetrahydrofolate ligase (EC 6.3.4.3) -
Clostridium thermaceticum g11144816|gb|AA23240.1|
(J02911) formyltetrahydrofolate synthetase (FTHFS) (ttg
start codon) (EC 6.3.4.3) [Moorella thermacetica], score
744, E-value 0"
/codon_start=1
/transl_table=11
/product="Formyltetrahydrofolate synthetase"
/protein_id="AAM25530.1"
/db_xref="GI:20517399"
/translation="MKSDEIAOEAQKMLHIDVAKIGIEEDYLEEYVKKAKISAL
WQIKRKRQKGLIVATIPTPAGEGKTTTGVGLGALAKLGRKAMIALREPLGSEF
GIRGAGGCGYSOVPMEDINLFTSDIRAITPAHNLAMIDNTHHNEINIDIRT
ITWKRAMDNDRLREITIVLGSKANGITPRQDITIKNLVOTIENVAEFGGFANA
LGDIIAAYDKDGPVTPARDLKAGAAVLLKQDIKPNLVOTIENVAEFGGFANA
HCGNSLARTYGLKADLVTEAGFGADGAEGAFVKAFFGGLPNAAVAFVAVL
KHGKGVKEDLOKEDVEAVRGIEENKEQVENRKFQVPAVVALNKFEVTEIEIEV
RRACDRIGVMAEAYWEKGGEGTILAEKIVIKADTPSNRFLYVNPVLIKRLIIT
AREITGADGVEYTAALKIDIANIEKLGDKMPVYAKTQYSISDDPKLGRPGFMT
VRELRTSRGAGTVALTGDTMTPGLPKHPAENIDIDENGRIKGLP"

misc_feature complement(4003..5667)
/gene="MIS1"
Query Match 2.18; Score 20; DB 1; Length 13610;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 828 GCAGAAAGGAGATGAGGCT 847
Db 11315 GCAGAAAGGAGATGAGGCT 11296
RESULT 50
LOCUS CEB0413 39726 bp DNA linear INV 02-SEP-1999
DEFINITION Caenorhabditis elegans cosmid B0413, complete sequence.
ACCESSION Z92824
VERSION 292824.1 GI:3873778
KEYWORDS Htg.
SOURCE
ORGANISM Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peleodierinae; Caenorhabditis.
1 (bases 1 to 39726)
Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Conson,A., Craxton,M., Connell,M., Copsey,T., Cooper,J.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kersey,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurry,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Riken,L., Roopra,A.,
Saunders,D., Showkeen,R., Smalton,N., Smith,A., Sonnenhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaubin,M.,
Vaughan,K., Waterson,R., Watson,A., Weinstock,L.,
Wilkinson-Sproat,J. and Wohldman,P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
PUBMED 7906398
REFERENCE 2 (bases 1 to 39726)
AUTHORS Basham,V.
TITLE Direct Submission
COMMENT Submitted (09-MAR-1997) Louis, MO 63110, USA. E-mail:
jesse.sanger.ac.uk or rtw@nemastode.wustl.edu
On Nov 16, 1998 this sequence version replaced gi:1877054.
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see: "
http://webcace.sanger.ac.uk/cgi-
bin/display?db=wormace&class=sequence&object=B0413
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
Location/Qualifiers
1..39726
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
join(3320..3396,3591..3712,4273..4322)
/gene="B0413.1"
join(3320..3396,3591..3712,4273..4322)
/gene="B0413.1"
/codon_start=1
/protein_id="CAB07306.1"
/db_xref="GI:4008332"


```
gene
/translation="MKFTFLVLITLTVLSCVSAEMDSXTPENXPKSNAPHEIIG
SMIGAGRHVRNKMSSQGXYSXSDPQOXXXDXDK"
complement(join(4445..4573,5010..5139,5190..5275,
6469..6587,6740..6846,7023..7184,7843..7973,8258..8480,
9088..9188))
/gene="B0413.2"
complement(join(4445..4573,5010..5139,5190..5275,
6469..6587,6740..6846,7023..7184,7843..7973,8258..8480,
9088..9188))
/gene="B0413.2"
/notes="similar to 4 transmembrane segments integral
membrane proteins"
/codon_start=1
/protein_id="CAB07307.1"
/db_xref="GI:4008333"
/translation="MSCVACVAVGKRYKSNXCSAOTRYRLKPFEPFLEFOVET
QCFEKRTKSLFLNKGKCTFRQFTGIRPLFLPDIQKLNSTFTFEKFKSL
QCKFEKXATSKIRGXVXXVXXVSPSYHNHGSVSDMTYTTDNTVDSEAN
RAGCGVMAVYGIETANIVFLVIGLXLMGWLKIDSRPNTISERYQAVOEAWEA
PTXFAVSYIXIVLGAVMVAMAGCCGIGRSRPLIYSMTVFLIVATLSGIIYL
KKDGETVPAPISEFEDRIKDPVIXKXDCGHRIMIALIGESVTIIVESAVLQOV
LITCFALYFVFMREKVKIIVEXPHRHSRLTRDLDLPLQIQKRRRASK
ER"
join(13685..13798,14087..14261,14308..14449,15299..15448,
15497..15760,16227..16332)
/gene="B0413.3"
join(13685..13798,14087..14261,14308..14449,15299..15448,
15497..15760,16227..16332)
/gene="B0413.3"
/notes="predicted using GeneFinder"
/codon_start=1
/protein_id="CAB07304.1"
/db_xref="GI:3873780"
/translation="MCORTIAPRLKCCANARBMHNAFASRYACXHSKXITISVT
LSTPKTKQKSELISPPAKPYMLAIGDYSGGCNPDIPTHSPAWLDDSCXS
SKAPFLVSQKTXATTSLSLJSCSGSTVDHGLISNGMKEFFVSHIXAVARLE
ENITRVLXLDYVDYTRNKKGVDSGAFGACEXXFXIXLIXXIIIFXKRIQKX
KGLXIRFELSRSPIQISLXKXKXFOKLNLRKAKXHGWTVDITYIFPTRGIXS
KSLIRFRESMRLOGNEGSPHXPEDAHXXCREXIEIG"
complement(join(16581..16640,16698..16822,17910..18332,
18748..18830,18902..19051,19138..19232,19383..19445,
19632..19732,20977..21018,21095..21248))
/gene="B0413.4"
complement(join(16581..16640,16698..16822,17910..18332,
18748..18830,18902..19051,19138..19232,19383..19445,
19632..19732,20977..21018,21095..21248))
/gene="B0413.4"
/codon_start=1
/protein_id="CAB07308.1"
/db_xref="GI:4008334"
/translation="MSDGNRRCFPDFLRFEGFVWKIMONTNYISXPLIFPSHLILI
XGCTYXAXASAKIYLSNCKNFKOCIDGRPLIDENLSTFSCORCGPDPFENOLC
PMTTASISLRPAAGKPNLXSRSTRKAVFOLSSXELRTLAXEYTPKLCNKDTS
OTIAKMYFDITLSCLPFGCCEDPDLKALRTGEEEBENLEXXAPKRXKLVIS
LHMSAFAGRSYPOAIDVIDIOLKXHIRGPXAIEPPDLVIVHKSDFRFXANLEX
SITVDEIXEKSENLAIEDNLLSTMEGATVPXTSNLILELDEXXGXSSLESL
EKSGNSGEXKXENRKRKXKXSEHSFEXPORNKSCCATVATRLCSCGASSOF
VYRMKEVGVCSPGYGCXKQONVAHPTRECBQXCD"
complement(19271..19342)
/gene="B0413.t2"
complement(19271..19342)
/gene="B0413.t2"
/notes="TTC Glu E-tRNA"
/product="tRNA-Glu"
/predicted using tRNAscan-SE 1.11
preliminary prediction
similar to tRNA-Glu"
join(28590..28669,29866..30559)
/gene="B0413.5"
join(28590..28669,29866..30559)
/gene="B0413.5"
/codon_start=1
/protein_id="CAB07309.1"
```

```
/db_xref="GI:4008335"
/translation="MILXHVYRHSFRTRRSHQKADFERFAXXPPLASSXGXSXXI
PSLSTFPLIGFEGVYTAIOSTFGMKNRIGEKIDKSDSGISDASEASDSECOIST
NFXOSISCKIMOKITTEMLKSIIDYIALCOKPLXLRXTXOIFRQAKXXVLR
IXKXKKKILIRISXNLKOPFTKMLGXSGXHLIFRCAKXIXNFENLENRXFX
XETKEIXAEXGEXXILERSAEXIIVNMGVWVWSX"
complement(join(32454..32734,33069..33211,33532..33701,
33775..33849,34037..34090))
/gene="B0413.6"
complement(join(32454..32734,33069..33211,33532..33701,
33775..33849,34037..34090))
/gene="B0413.6"
/notes="predicted using GeneFinder"
/codon_start=1
/protein_id="CAB07305.1"
/db_xref="GI:3873783"
/translation="MTNOKVITREXIHRSKSDVAFVHILRGKPYTKPLPLPTVQEA
DEPKXPYXSNHREPLRLSTSDQWDTNNRNRSSMTRLSRFOPMDIPTXGP
LAXXXIRRSKIRLREISELRMRIDRMSLSTDOMXOPFRQHKYXRLNLEMAAX
QIDHLEQARAHLSRIILTEPTPTPEERKIDISWTENHEKLLAEMRAOAPPTHTV
FOEXPTQIVVPPVXGXXLS"
complement(36074..36146)
/gene="B0413.t1"
complement(36074..36146)
/gene="B0413.t1"
/product="tRNA-Phe"
/notes="GAA Phe F-tRNA"
predicted using tRNAscan-SE 1.11
preliminary prediction
similar to tRNA-Phe"
join(37289..37544,39351..39461,39550..39627,39703..39726)
/gene="B0413.7"
join(37289..37544,39351..39461,39550..39627,39703..39726)
/gene="B0413.7"
/notes="cDNA EST yk481h5.5 comes from this gene
cDNA EST yk246b6.5 comes from this gene
cDNA EST EMBL:CO9106 comes from this gene
cDNA EST yk339g8.5 comes from this gene
cDNA EST yk440g8.5 comes from this gene"
/codon_start=1
/protein_id="CAB07310.1"
/db_xref="GI:4008336"
/translation="MESTOKDIKLEPDEFNYITAPGAPESNLAENPENPVLSKXPL
EKRPAPFSSKRSIYWEHXITWVAESQRAACNCAEFQVKGSTSLRLHRLKTIHPAA
INDPENPPVARRKXKMEXYTIVIEDTKSYMXNLDELFINWDSAPRGGLPD"
BASE COUNT 12836 a 6244 c 6214 g 12719 t 1713 others
ORIGIN
```

Search completed: April 26, 2003, 11:55:49
Job time : 3481 secs

Query Match 2.1%; Score 20; DB 3; Length 39726;
Best local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 385 CTATTTTGAATTCGAA 404
|||||
DB 24283 CTATTTTGAATTCGAA 24302

THIS PAGE BLANK (USPTO)